AD_____

Award Number: DAMD17-03-1-0034

TITLE: Short- and Long-Term Effects in Prostate Cancer Survival: Analysis of Treatment Efficacy and Risk Prediction

PRINCIPAL INVESTIGATOR: Alexander Tsodikov, Ph.D.

CONTRACTING ORGANIZATION: University of California, Davis
Davis, California 95616-8671

REPORT DATE: March 2005

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 2005	Annual (1 Mar 04 - 28 Feb 05)

4. TITLE AND SUBTITLE
Short- and Long-Term Effects in Prostate Cancer Survival: Analysis of Treatment Efficacy and Risk Prediction

5. FUNDING NUMBERS
DAMD17-03-1-0034

6. AUTHOR(S)
Alexander Tsodikov, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Davis
Davis, California 95616-8671

E-Mail: atsodikov@ucdavis.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

We have completed methodology and software development for point and interval estimation and variable selection for compound Nonlinear Transformation Models. We have built a number of candidate compound models for prostate cancer and verified their properties analytically and by simulations. Finally, we used the new software and methodology to apply these models to a number of real and simulated test data sets.

14. SUBJECT TERMS
No subject terms provided.

15. NUMBER OF PAGES
70

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Contents

1 Statement of Work

Short- and Long-term Effects in Prostate Cancer Survival: Analysis of Treatment Efficacy and Risk Prediction
Alexander Tsodikov, Ph.D.

There has been no change in the scope of work. In order to compensate for the period of inactivity associated with the change of the PI institution last year, a no cost extension of one year has been granted. The payment schedule and percent effort of the personnel on the grant was modified accordingly and payments have been spread equally over the two years starting with the grant transfer. A breakdown below shows what has been accomplished in the first two years of the project.

Tasks accomplished in the first 8 Months of the project at the University of Utah

Task 1. Develop model-building techniques

Task 2A. Develop estimation and hypothesis testing

(a) Develop point estimation

(b) Develop simulation algorithms

(c) Develop hypothesis testing

Tasks completed at the University of California at Davis during year 2

Task 2B. Develop estimation and hypothesis testing

(d) Develop software implementation

(e) Study models and methods by simulation

Task 3. Develop variable selection procedures

Task 4A. Preliminary analysis of the data for significant effects

(a) Apply estimation, hypothesis testing and variable selection to a test subset of SEER data.

Tasks to be completed at the University of California at Davis during year 3

Task 4B. Continue analysis of the data for significant effects

(a) Continue application of estimation, hypothesis testing and variable selection to SEER and MSKCC data.

(b) Identify a model for prostate cancer biochemical recurrence, prostate cancer specific survival, and overall survival using methodology and software developed in Tasks 1-4.

Task 5. Computer-intensive approaches to prognosis and validation

2 Objectives

There has been no change in the project objectives. The specific aims of this project are

1. To provide a statistical model that reproduces the complex survival responses in prostate cancer.
2. To develop methodology for analysis of prognosis after treatment for prostate cancer taking into account the long- and short-term effects of prognostic factors and treatment.
3. To develop statistical software implementing model-building, estimation, construction of prognostic indices, conditional survival prognosis, and assessment of the quality of prognostic classifications based on the new models.
4. To apply the models and methodology to analyze post-treatment survival of patients with prostate cancer using data from the Memorial Sloan Kettering Cancer Center and the SEER database.

3 Introduction

The goal of this proposal is to investigate a novel approach to the analysis of post-treatment survival of prostate cancer patients: the decomposition of the diversity of survival patterns into short-term and long-term effects. We proposed to identify a model of prostate cancer survival incorporating long- and short-term effects of prognostic factors and treatment. Novel statistical tools are being developed to make such models work for better prognosis of prostate cancer patients. Year 1 at the University of Utah was primarily devoted to development of methodology for point estimation and hypothesis testing. While continuing methodological research in Year 2, we focused on the delivery aspect of the progect addressing software development and implementation of the algorithms, testing them by simulations, development of tools for multivariate analysis and variable selection and preliminary applications of these tools to real data.

4 Nonlinear Transformation Models

Definition 4.1 *Let $\gamma(x \mid \beta, z)$ be a parametrically specified distribution function with x-domain being the interval $[0,1]$. Let $F(t)$ be a nonparametrically specified baseline survival function. A semiparametric regression survival model is called a Nonlinear Transformation Model if, conditional on the covariates z, its survival function G can be represented in the form*

$$G(t \mid \beta, z) = \gamma(F(t) \mid \beta, z). \tag{1}$$

The function γ is called the NTM-generating function by analogy with the probability generating functions.

Note that $F(t) = \exp(-H(t))$ where $H(t)$ is the baseline cumulative hazard function.

The class of NTM was developed in Year 1 of the project. In the Year 1 report we proposed the Quasi-EM (QEM) algorithm for ML estimation. The algorithm is based on imputation of the predictor in a Nelson-Aalen-Breslow-like estimator using the posterior risk function

$$\Theta(x \mid \cdot, c) = c + x\frac{\gamma^{(c+1)}(x \mid \cdot)}{\gamma^{(c)}(x \mid \cdot)}, \tag{2}$$

where $\gamma^{(c)}(x\,|\,\cdot) = \partial^c\gamma(x\,|\,\cdot)/\partial x^c$, $c = 0, 1, \ldots$, $\gamma^{(0)}(x\,|\,\cdot) = \gamma(x\,|\,\cdot)$. The key requirement that ensures monotonicity, convergence and an EM-like behavior of the QEM algorithm is that the function $\Theta(x\,|\,\cdot, c)$ is nondecreasing in x.

Let t_i, $i = 1, \ldots, n$ be a set of failure times, arranged in increasing order, $t_{n+1} := \infty$. Associated with each t_i is a set of subjects $\mathcal{D}_i$ with covariates z_{ij}, $j \in \mathcal{D}_i$ who fail at t_i, and a set of subjects $\mathcal{C}_i$ with covariates z_{ij}, $j \in \mathcal{C}_i$ who are censored at time $t \in [t_i, t_{i+1})$. The observed event for the subject ij is a triple (t_i, z_{ij}, c_{ij}), where c is a censoring indicator, $c = 1$ if failure, $c = 0$ if right censored. Let H be the baseline cumulative hazard, with $H(0) = 0$. We assume than $H(t)$ is a step function with jumps at the failure times t_i, $i = 1, \ldots, n$. As a step-function, H can be characterized by the vector $h = (h_1, ..., h_n)$, where h_i is the jump at t_i. With this notation, under an NT model and non-informative censoring, the likelihood of survival data takes the form

$$\ell = \sum_{i=1}^{n} D_i \log(h_i) + \sum_{i=1}^{n} \sum_{j \in \mathcal{C}_i \cup \mathcal{D}_i} \log \vartheta(F_i\,|\,\beta, z_{ij}, c_{ij}), \tag{3}$$

where

$$\vartheta(x\,|\,\cdot, c) = x^c\gamma^{(c)}(x\,|\,\cdot),$$

D_i is the number of failures associated with t_i and

$$F_i = F(t_i) = \exp(-\sum_{l=1}^{i} h_l).$$

Differentiating ℓ with respect to h and setting the score equal to 0 we obtain $\hat{h}(\beta)$ as the solution of the functional self-consistency equation

$$\hat{h}_m = \frac{D_m}{\sum_{(i,j)\in\mathcal{R}_m} \Theta(F_i\,|\,\beta, z_{ij}, c_{ij})}, \quad m = 1, \ldots, n, \tag{4}$$

where F_i is a function of $h_1, \ldots, h_i$, Θ is given by (2) and $\mathcal{R}_m$ is the set of subjects at risk just prior to t_m, $\mathcal{R}_m = \{(i,j) : i \geq m,\ j \in \mathcal{C}_i \cup \mathcal{D}_i\}$.

Solving the self-consistency equation by iterations in h (the QEM procedure), we obtain its solution as a fixed-point of a contracting operator.

5 Hypotheses Testing

5.1 Existing methods

In semiparametric models considered in this project, the parameter is partitioned as (β, H) with β a low-dimensional parameter representing regression coefficients, and H a high-dimensional nuisance parameter, representing the baseline cumulative hazard.

Consider the profile likelihood

$$\ell_{pr}(\beta) = \max_{H} \ell(\beta, H).$$

The observed profile information matrix will be denoted I_{pr},

$$I_{pr} = -\left.\frac{\partial^2 \ell_{pr}(\beta)}{\partial\beta\partial\beta^{\mathrm{T}}}\right|_{\beta=\hat{\beta}}.$$

We compare our method to the following three existing techniques used to estimate the profile information matrix that amount to particular forms of numerical differentiation of the second order.

1. *Discretized second derivative.* Corollary 3 of [Murphy and van der Vaart, 2000] shows that under certain conditions
$$-2\frac{\log \ell_{pr}(\hat{\beta}+h_n v_n)-\log \ell_{pr}(\hat{\beta})}{nh_n^2} \to v^{\mathrm{T}} I_{pr} v,$$
for all sequences $v_n \to v \in \mathbb{R}^d$ and $h_n \to 0$ such that $(\sqrt{n}h_n)^{-1} = O_P(1)$.

   This result can be used to derive an estimate of I_{pr}. We use its deterministic version with $v_n \equiv v = a_1 e_i + a_2 e_j$, $1 \le i \le j \le d$, where e_i are Euclidean basis vectors and $a_1, a_2 = 0$ or 1.

2. *Fitting a Quadratic Form.* In many cases the profile likelihood surface around the true β is asymptotically quadratic. Nielsen et al. [1992] proposed fitting a quadratic form to $\ell_{pr}(\beta)$ in some domain around the maximum likelihood estimator, $\hat{\beta}$, and to derive an approximate profile information matrix using the estimated coefficients of the form. Specifically, let $\Delta\beta$ be a vector of deviations of the β values sampled in the vicinity of $\hat{\beta}$, and let $\Delta\ell_{pr}$ be the induced vector of deviations of the profile likelihood from its maximum value, $\ell_{pr}(\hat{\beta})$. Then, if $\Delta\beta$ is sufficiently small
$$\Delta\ell_{pr} \approx \frac{1}{2}\Delta\beta^{\mathrm{T}} I_{pr} \Delta\beta.$$
   Fitting the quadratic form $(1/2)\Delta\beta^{\mathrm{T}} A \Delta\beta$ to points $(\Delta\beta, \Delta\ell_{pr})$ by least squares produces an estimate, $\hat{A}$, of the profile information matrix I_{pr}.

3. *Numerical Differentiation of the Profile Likelihood.* Standard numerical algorithms can be used to numerically differentiate a function. The procedure usually involves evaluating the target function at some pre-specified knots and interpolating the surface. The interpolating function can then be differentiated to estimate the curvature of the surface. We use Ridder's method [Press et al., 1994] in the examples presented in Section 5.3.

Globally the likelihood surface is not quadratic. The quadratic approach has the difficulty that a sufficiently small domain around $\hat{\beta}$ where the likelihood surface can be well approximated by a quadratic form is not well defined. It our implementation of this method we limit the domain to points that are not rejected at 0.05 significance level by the LR test (applied informally). Numerical errors with the quadratic method often lead to estimates of the profile information matrix that are not positive definite, particularly if the number of covariates is large.

5.2 The new exact method

Denote by h a vector representing a set of jumps of the cumulative hazard function $h_i = H(t_i) - H(t_i - 0)$.

Implicit differentiation of the profile likelihood yields the following expression for the profile information matrix

$$I_{pr} = I_{\beta\beta} + \hat{h}_\beta^{\mathrm{T}} I_{hh} \hat{h}_\beta + \hat{h}_\beta^{\mathrm{T}} I_{h\beta} + I_{h\beta}^{\mathrm{T}} \hat{h}_\beta, \tag{5}$$

where
$$\hat{h}_\beta = \left.\frac{\partial \hat{h}}{\partial \beta}\right|_{\beta=\hat{\beta}} \quad \text{and} \quad I_{ab} = -\left.\frac{\partial^2 \ell(\beta, h)}{\partial a \partial b^{\mathrm{T}}}\right|_{(\hat{\beta},\hat{h})}$$
with a and b equal to β or h.

Notice that I_{pr} has dimension $d \times d$, $d = \dim(\beta)$. Therefore only a small matrix needs to be inverted in order to get an estimator of the covariance matrix of regression coefficients.

The difficulty in (5) is that since $\hat{h}(\beta)$ is defined implicitly, so is the potentially large Jacobian matrix $\partial\hat{h}/\partial\beta$. Therefore, the Jacobian is generally unavailable in a closed form. The success in the calculation of the profile information matrix is determined by the existence of an efficient numerical method to compute $\partial\hat{h}/\partial\beta$.

For Nonlinear Transformation Models considered in this project, $\partial\hat{h}/\partial\beta$ can be obtained by solving a system of linear equations with a special structure. This specific structure of the linear system can be exploited to derive an efficient numerical solution given in Proposition 5.1.

First we obtained $I_{\beta\beta}$, $I_{h\beta}$ and I_{hh}.

The H–score of an NT model is,

$$\frac{\partial\ell}{\partial h_k} = \frac{D_k}{h_k} - \sum_{(i,j)\in\mathcal{R}_m} \Theta(F_i \,|\, \beta, z_{ij}, c_{ij}). \tag{6}$$

Differentiating the H–score with respect to β we get,

$$-\frac{\partial\ell^2}{\partial h_k \partial\beta_m} = \sum_{(i,j)\in\mathcal{R}_k} \frac{\partial\Theta}{\partial\beta_m}(F_i \,|\, \beta, z_{ij}, c_{ij}). \tag{7}$$

Evaluation of derivatives of Θ or γ with respect to β depends on the parameterization of the model's predictor as a function of explanatory variables z, which is model–specific. Once a model is specified, the calculation of $I_{\beta\beta}$ and $I_{h\beta}$ is straightforward.

Since $F_i = \exp(-\sum_{l=1}^{i} h_l)$, we have

$$\frac{\partial\Theta(F_i \,|\, \cdot)}{\partial h_m} = \begin{cases} Q(F_i \,|\, \cdot), & m \le i, \\ 0, & m > i, \end{cases} \tag{8}$$

where

$$Q(x \,|\, \cdot, c) = -x\frac{\partial\Theta(x \,|\, \cdot, c)}{\partial x} = -(\Theta(x \,|\, \cdot, c) - c)(\Theta(x \,|\, \cdot, c+1) - \Theta(x \,|\, \cdot, c)). \tag{9}$$

Note that $\partial\Theta(F_i \,|\, \cdot)/\partial h_m$ is a constant in m for $m \le i$ or $m > i$.

From (8) it follows that,

$$-\frac{\partial^2\ell}{\partial h_k \partial h_m} = \sum_{(i,j)\in\mathcal{R}_{\max\{k,m\}}} Q(F_i \,|\, \beta, z_{ij}, c_{ij}) + \frac{D_k}{h_k^2} 1_{\{k=m\}}, \tag{10}$$

where

$$1_{\{k=m\}} = \begin{cases} 1, & k = m, \\ 0, & k \ne m. \end{cases}$$

From this we get I_{hh}.

Now we turn our attention to the Jacobian $\partial\hat{h}/\partial\beta$. Proposition 5.1 gives the main result used to efficiently calculate $\partial\hat{h}/\partial\beta$ in the case of NT models.

Proposition 5.1 *Let D be an $n \times n$ diagonal matrix with diagonal elements $d_i \ne 0$, $i = 1, \ldots, n$. Let $R = (R_{kl})$ be an $n \times n$ matrix, with $R_{kl} = \sum_{i=\max\{k,l\}}^{n} a_i$, where a_i, $i = 1, \ldots n$ are real numbers. Let b be an n-dimensional vector.*

Define the functions $\varphi_k : \mathbb{R} \to \mathbb{R}$, $k = 1, \ldots, n$ recursively as

$$\begin{aligned} \varphi_n(y) &= \frac{b_n}{d_n} - \frac{a_n}{d_n} y, \\ \varphi_k(y) &= \frac{1}{d_k}\left(b_k - \sum_{i=k}^{n} a_i y + \sum_{l=k+1}^{n} \sum_{i=k}^{l-1} a_i \varphi_l(y) \right), \quad k = n-1, \ldots, 1, \end{aligned}$$

for y in $\mathbb{R}$. Let $\tilde{\varphi}:\mathbb{R}\to\mathbb{R}$ be the function given by $\tilde{\varphi}(y)=\sum_{k=1}^{n}\varphi_k(y)$ and let

$$\tilde{y}=\frac{\tilde{\varphi}(0)}{1+\tilde{\varphi}(0)-\tilde{\varphi}(1)}.$$

Then the solution to the system of equations $(D+R)x=b$ is the n-dimensional vector $x=(\varphi_1(\tilde{y}),\ldots,\varphi_n(\tilde{y}))^T$.

We now show that the Jacobian $\partial\hat{h}/\partial\beta$ satisfies a relationship of the form as discussed in Proposition 5.1. Differentiating the self-consistency equation (4) implicitly, we get that $\hat{h}$ satisfies the relationship

$$\frac{\partial\hat{h}_m}{\partial\beta_k}=-\frac{\hat{h}_m^2}{D_m}\left(\sum_{l=1}^{n}\sum_{(i,j)\in\mathcal{R}_{\max\{m,l\}}}Q(F_i\,|\,\beta,z_{ij},c_{ij})\frac{\partial\hat{h}_l}{\partial\beta_k}+\sum_{(i,j)\in\mathcal{R}_m}\frac{\partial\Theta}{\partial\beta_k}(F_i\,|\,\beta,z_{ij},c_{ij})\right), \tag{11}$$

where Q is the function given in (9).

Let D be the diagonal matrix with elements

$$d_m=\frac{D_m}{(\hat{h}_m)^2},\quad m=1,\ldots,d.$$

Let $R=(R_{ml})$ with $R_{ml}=\sum_{i=\max\{m,l\}}^{n}a_i$, where

$$a_i=\sum_{j\in\mathcal{C}_i\cup\mathcal{D}_i}Q(F_i\,|\,\beta,z_{ij},c_{ij}),\quad i=1,\ldots,n$$

and for $k=1,\ldots,d$ let

$$b^{(k)}=\left(-\sum_{(i,j)\in\mathcal{R}_1}\frac{\partial\Theta(F_i\,|\,\beta,z_{ij},c_{ij})}{\partial\beta_k},\ldots,-\sum_{(i,j)\in\mathcal{R}_n}\frac{\partial\Theta(F_i\,|\,\beta,z_{ij},c_{ij})}{\partial\beta_k}\right)^{\mathrm{T}}.$$

It follows from (11) that

$$\frac{\partial\hat{h}}{\partial\beta_k}=-D^{-1}\left(R\frac{\partial\hat{h}}{\partial\beta_k}-b^{(k)}\right).$$

Hence,

$$(R+D)\frac{\partial\hat{h}}{\partial\beta_k}=b^{(k)}.$$

Therefore, for each $k=1,\ldots,n$ the vector $\partial\hat{h}/\beta_k$ can be obtained from Proposition 5.1. We now have all the components of (5) defined. This completes the exposition of our method.

5.3 Comparison of methods

We compared the performance of four methods to compute the observed profile information matrix:

1. *Discretized.* The estimation is based on the result of Corollary 3 in Murphy and van der Vaart [2000].

2. *Quadratic.* This approach approximates the profile likelihood surface by a quadratic form and derives the estimate of the information matrix from the coefficients of the form fitted to the surface (Nielsen et al. [1992]).

3. *Numerical.* The calculation of the observed profile information matrix is carried on using Ridder's numerical differentiation of the profile likelihood function (Press et al. [1994]).

4. *Exact.* This is our new method.

PO model was used as a basis for all our comparisons. The validity of NPMLE and the profile likelihood for this model has been demonstrated elsewhere.

5.3.1 Real data

We continue the example considered in Year 1 report. We use data from the National Cancer Institutes Surveillance Epidemiology and End Results (SEER) program. Using the publicly available SEER database, 11621 cases of primary prostate cancer diagnosed between 1988 and 1999. Two groups of patients representing stage at diagnosis of the disease are considered, hence the predictor in the PO model has a single parameter β. The log odds ratio β measures the disadvantage of being in the distant stage relative to local/regional stage. The QEM algorithm was applied to fit the PO model to the data. The maximum likelihood estimate of β is $\hat{\beta} = -3.251$. Confidence intervals for β were obtained using the Wald statistic based on the profile information matrix. The confidence interval based on the quadratic approximation of the profile information matrix is $(-3.416, -3.086)$ and the one obtained through the exact profile information matrix is $(-3.415, -3.086)$. Excellent concordance of the two confidence intervals is due to the large sample size and the small dimension of the regression parameter, a situation when approximating methods tend to be accurate.

In the case of a single parameter, the observed profile information matrix is a scalar. The estimates of the observed profile information matrix are 142.1011, 141.2158 and 141.7424 for the Discretized, Quadratic and Numerical approaches respectively and the Exact value is 141.7423. Although the values are quite similar it is clear that the discretized and quadratic approaches depart from the true value.

5.3.2 Simulations

Simulations were performed using a parametric PO model where the baseline survival function F was specified according to a Weibull distribution. In a set of experiments, samples of size ranging between 100 to 1000 were generated from the Weibull PO model with one continuous covariate uniformly distributed on [-1,1], with regression coefficient $\beta_3 = 3.45$ and one categorical covariate (Group) with 3 levels. Simple contrast was used to code for the levels of the Group with regression coefficients of $\beta_1 = 1.3$ (Group 2 vs. Group 1), and $\beta_2 = 2.4$ (Group 3 vs. Group 1). The baseline survival function F was generated from a Weibull distribution with shape parameter 2 and median of 1.33. Censoring was generated from a Weibull distribution with both shape and median equal to 1. To assess the speed of performance of the three methods we calculated the number of operations required to compute the exact information matrix and its approximations. Evaluation of Θ, γ, their analytically specified derivatives or similar comparable procedures were counted as one operation. Figure 1 shows the number of operations by sample size and method. The estimation algorithms were calibrated so that the relative error of the three methods (Discretized, Quadratic, Numerical) was about the same. Regardless of the sample size, the exact calculation outperformed the approximate methods. Inference based on the discretized second derivative requires between 10 and 30 times as many operations as the exact calculation. The quadratic approach requires between 60 and 200 times as many operations as the calculation of the exact I_{pr} matrix. The numerical method is computationally very costly requiring between 600 and 7000 as many operations as the exact approach. However, the numerical approach behaves better than the

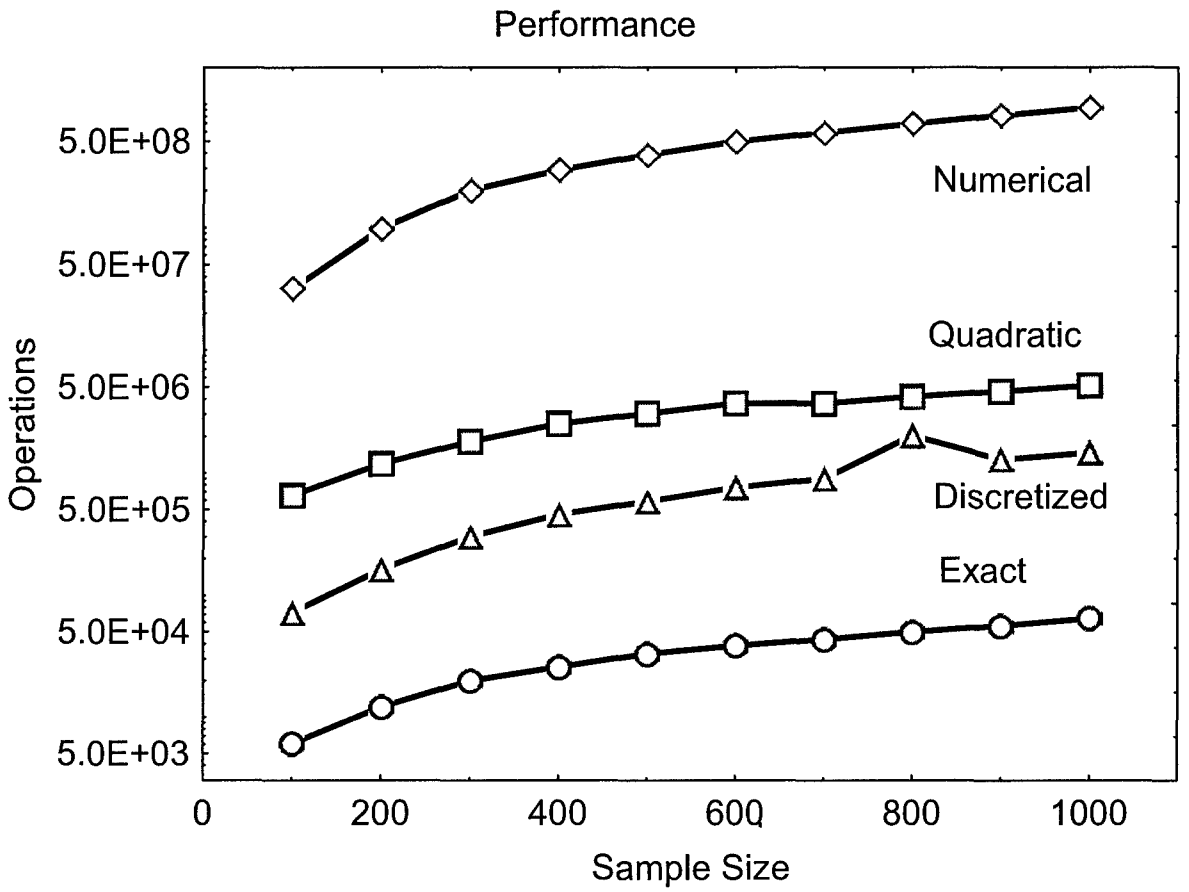


Figure 1: Operations by sample size characteristics of three methods of computation of the observed profile information matrix. The Exact method developed in this paper shows the highest numerical efficiency.

other two methods in terms of relative error. A sample of size 500 was used to find the smallest possible relative error of the method when adjusting the different parameters involved. The best relative error achieved by the Discretized method was 0.01 and $8.13\ 10^5$ operations were required. This number was 0.013 for the quadratic approach with $5.32\ 10^6$ operations required, while the numeric approach achieved a relative error of $8\ 10^{-7}$ and required $3.87\ 10^8$ operations.

As the Exact method makes no compromise and delivers the exact numerically efficient solution to the problem for the class of semiparametric Nonlinear Transformation Models, there is little point in using other alternative procedures with such models.

6 The meaning of imputation operator Θ

Consider a PH mixture model

$$G(t \mid \beta, z) = \mathrm{E}\left\{ F(t)^{U(\beta,z)} \,\middle|\, z \right\}, \tag{12}$$

where U is the frailty random variable. Suppose, we have an observation (t, z, c) sampled from the PH mixture model under independent censoring, where t is an observed survival time and c is a censoring indicator ($c = 0$ if t is a censored survival time, and $c = 1$ if t is a failure). Then, under

the PH mixture model (12), the conditional expectation of U, given the observed event $(t, \boldsymbol{z}, c)$ is given by

$$\mathrm{E}\{U(\cdot) \mid t, \cdot, c\} = (\Theta \circ F)(t \mid \cdot, c) = \Theta\left[F(t) \mid \cdot, c\right],$$

where the function Θ is given by (2). For brevity, we use $(\cdot)$ to suppress covariates and regression coeffitients $\boldsymbol{\beta}, \boldsymbol{z}$. While Θ is defined for NTMs, we also consider the probability generating functions subclass of γs associated with the PH mixture model as a motivation and to better understand the conditions that make the NTM-QEM tandem work.

Cauchy-Schwartz inequality can be used to show that for any PH mixture model, $\Theta\left[x \mid \cdot, c\right]$ is nondecreasing in x for any $c = 0, 1$. The nondecreasing character of the function Θ in the above statement is quite natural. The longer the subject stays event–free, the lower the subject's posterior risk, represented by Θ. So $\Theta\{F(t) \mid \cdot, c\}$ must be a nonincreasing function of t for both failure ($c = 1$) and censoring ($c = 0$) events. Since the survival function $F(t)$ is nonincreasing in t, $\Theta(x \mid \cdot, c)$ must be nondecreasing in x. It is interesting to note that the population hazard function for a heterogeneous population under the PH mixture model is expressed as $\lambda(t \mid \boldsymbol{z}) = \Theta\{F(t) \mid \cdot, 0\}h(t)$, where h is the hazard function corresponding to F. Even if $h(t)$ is increasing, the observed population hazard function may be a decreasing one through the decreasing behavior of $\Theta\{F(t)|\cdot, 0\}$ with time. This observation represents a selection effect of the risk set becoming "healthier" with time, as frail individuals leave the population. This effect was discovered and extensively studied in demography [Vaupel et al., 1979] in the context of misinterpretation of mortality trends.

With γ representing a PH mixture model, kth moments of the mixing variable U, $k = 1, 2, \ldots$, can be obtained through derivatives $\gamma^{(k)}$. Both Θ and QEM are defined using the derivatives up to second order of γ, $k = 1, 2$. Based on the above observations, NTM-QEM tandem is defined to follow second-order properties of the Frailty-EM frame. This is all that is needed to ensure the EM-like behavior of the QEM, and existence of all derivatives of γ (still a weaker assumption than that of a frailty model) is excessive for purposes of statistical inference.

As discussed in [Tsodikov, 2003], the property of non-decreasing Θ represents a generalized form of Jensen inequality on the primitive class of functions necessary to handle the QEM algorithm.

In addition to being a non-increasing function of time, the posterior risk $\mathrm{E}\{U(\cdot) \mid t, \cdot, c\}$ for PH mixture models ($\gamma \in \mathcal{P}$) has the following two natural properties.

1. Other things equal, the posterior risk of a failure is at least as high as a posterior risk of a censored subject $\mathrm{E}\{U(\cdot) \mid t, \cdot, 1\} \geq \mathrm{E}\{U(\cdot) \mid t, \cdot, 0\}$. This statement is valid in the general NTM form (see proposition below).

2. Since a censored observation at time $t = 0$ does not contribute any information on the risk, posterior risk for $t = 0$, $c = 0$ is the same as prior risk $\mathrm{E}\{U(\cdot)\}$. Expressing the mean of U through its p.g.f. $\gamma \in \mathcal{P}$, we have $\mathrm{E}\{U(\cdot) \mid 0, \cdot, 0\} = \mathrm{E}\{U\} = \gamma'(1 \mid \cdot)$.

Proposition 6.1 *Surrogate of posterior risk for NTM.*
Let $\Theta(x \mid \cdot, c)$, be the function defined by (2) and induced by some NTM generating function γ, given an event $(t, \cdot, c)$ observed on a subject. Then

(A) If $\Theta(x \mid \cdot)$ is a non-decreasing function of x, then

$$\Theta(F(t) \mid \cdot, 1) \geq \Theta(F(t) \mid \cdot, 0) > 0 \tag{13}$$

(B) If $\gamma \in \mathcal{P}$ is a p.g.f. of some nonnegative random variable U, then

$$E\{U \mid t, \cdot, 1)\} \geq E\{U \mid t, \cdot, 0)\} > 0 \tag{14}$$

$$E\{U \mid 0, \cdot, 0)\} = E\{U \mid \cdot\} = \gamma'(1 \mid \cdot) \tag{15}$$

The graph of typical behavior of the posterior risk is given in Figure 2 based on the real data example considered earlier in this report.

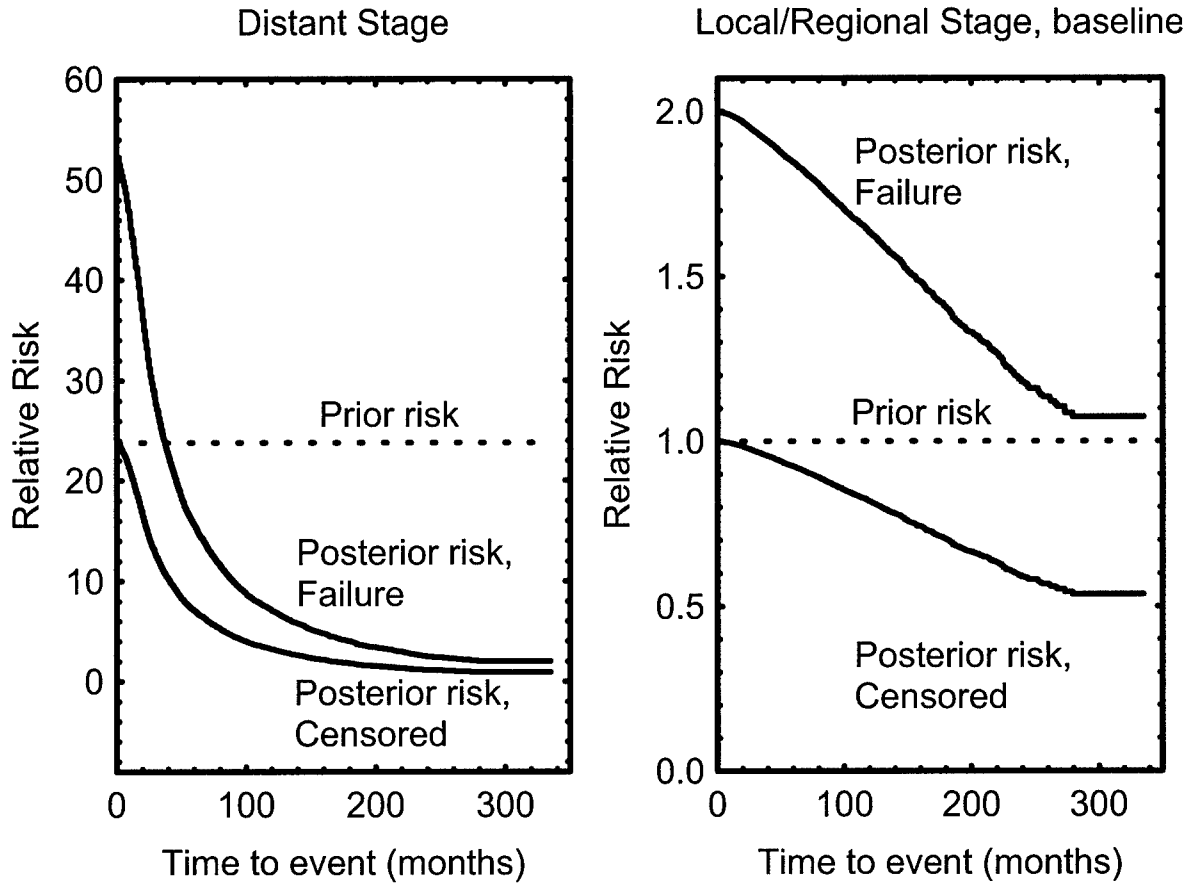


Figure 2: Posterior risk $\Theta(F(t) \mid \boldsymbol{\beta}, z, c)$ as a function of time to event t by type of event (failure, $c = 0$ and censoring $c = 1$), and Stage (z) (Local/Regional and Distant)

7 Compound models

In the Year 1 of the project, we developed a composition device for constructing hierarchical Nonlinear Transformation Models compatible with the QEM estimation framework. In this section we put this device into practice and show how is can be used to build new models that combine the features of simpler submodels.

7.1 PHPH Cure Model

This model extends the Improper PH model by introducing a PH short-term effect on the normalized baseline cumulative hazard $F \to F^{\eta(\boldsymbol{\beta}, \boldsymbol{z})}$,

$$G(t \mid \boldsymbol{\beta}, \boldsymbol{z}) = \exp\left\{-\theta(\boldsymbol{\beta}, \boldsymbol{z})[1 - F(t)^{\eta(\boldsymbol{\beta}, \boldsymbol{z})}]\right\}. \tag{16}$$

Here we note that the model is constructed by composition of NTM generating functions for the Improper PH model $\gamma_\theta(x) = e^{-\theta(1-x)}$ and the Proper PH model $\gamma_\eta(x) = x^\eta$,

$$\gamma_{\theta,\eta}(x\,|\cdot) = \gamma_\theta(x\,|\cdot) \circ \gamma_\eta(x\,|\cdot) =$$

$$[\exp\{-\theta(\cdot)(1-x)\}] \circ \left[x^{\eta(\cdot)}\right] = \exp\left\{-\theta(\cdot)\left(1-x^{\eta(\cdot)}\right)\right\}. \tag{17}$$

A review and history of this model is presented in [Tsodikov et al., 2003].

Note that γ_θ is a p.g.f. of a Poisson random variable, and γ_η is a p.g.f. of a nonrandom variable. Therefore the composition is a particular case of Aalen's device [Aalen, 1992]

$$U(\boldsymbol{\beta}, \boldsymbol{z}) = \sum_{k=1}^{\nu(\boldsymbol{\beta}_\theta, \boldsymbol{z})} \xi_k(\boldsymbol{\beta}_\eta, \boldsymbol{z}), \quad \sum_1^0 = 0, \tag{18}$$

with ν being Poisson(θ), and $\xi = \eta$ being nonrandom.

The chain rule developed in Year 1 immediately leads to

$$\Theta(x|\cdot, c) = \theta(\cdot)\eta(\cdot)x^{\eta(\cdot)} + c\eta(\cdot). \tag{19}$$

7.2 Γ-frailty model

Now, consider a model composed of the PH and the PO models.

The Γ–frailty model can be built as a composition of the NTM–generating functions corresponding to the PO and the proper PH models. As a result of the composition $\gamma = \gamma_\theta \circ \gamma_\eta$, we have

$$G\{t|\,\boldsymbol{\theta}(\cdot), \boldsymbol{\eta}(\cdot)\} = \left\{\frac{\theta(\cdot)}{\theta(\cdot) + H(t)}\right\}^{\eta(\cdot)}. \tag{20}$$

Indeed,

$$\gamma_{\theta,\eta}(e^{-s}|\cdot) = \left[\frac{\theta(\cdot)}{\theta(\cdot)+s}\right]^{\eta(\cdot)}$$

is the Laplace transform of a Γ-distribution with scale parameter θ and shape parameter η, and we have the interpretation of the compound model (20) as a Γ–frailty model.

Note that since an exponentially distributed random variable corresponding to γ_θ is a continuous one, the above composition is not a particular case of (18).

The compound Θ is derived from the chain rule is

$$\Theta(x\,|\,\cdot, c) = \frac{\eta(\cdot) + c}{\theta(\cdot) - \log x}. \tag{21}$$

7.3 Score test for long- and short-term effects

Keeping in mind the challenge of computer intensive regression and prediction approaches to the analysis of large sample prostate cancer data to be undertaken in Year 3, we addressed an alternative strategy of two-sample testing based on the partial likelihood in a time-dependent PH model [Broët et al., 2004]. The two-sample statistics are suited for testing equality of survival functions against improper semi-parametric accelerated failure time alternatives. These tests are designed for comparing either the short- or the long-term e.ect of a prognostic factor, or both, and are thus based on a model conceptually similar to PHPH. The model was motivated by the Weibull distribution,

$$G(t\,|\,\boldsymbol{\beta}, \boldsymbol{z}) = \exp\left\{-\theta(\boldsymbol{\beta}, \boldsymbol{z})[1 - \exp\left\{-H(t)^{\eta(\boldsymbol{\beta}, \boldsymbol{z})}\right\}]\right\}, \tag{22}$$

where H is a baseline cumulative hazard. The proposed tests can be easily implemented using widely available software. This strategy may be used in regression tree methods where a fast two-sample test is needed that would take long- and short-term effects into account. A breast cancer clinical trial is presented as an example to demonstrate the utility of the proposed tests.

Generally the idea is to construct a compound model where one of the submodels for the long-term effect is the PH model. Let β_1, β_2 be the two regression coefficients modeling long- and short-term effects, respectively. For the score test, $\beta. \to 0$. The cumulative hazard for the compound model is expanded into a Taylor series, and linear terms in β_1 and two are kept in the derivation of the test statistic. Since the PH model is a long-term effect submodel, this expansion leads to $\beta_1 + \beta_2 w(t)$, where w is some non-decreasing function modelled nonparametrically. For the test for short-term effect and the test for homogeneity, the model under the Null hypothesis is a PH, and the score test is based on well known estimated for the PH model. The score two-sample statistics has asymptotic χ^2 distribution with one (test for short-term effect) or two (test for homogeneity) degrees of freedom.

The test for long-term effect represents the most difficult case, as the model under the Null hypothesis is not PH. Efficient estimators for β_2 and w need to be derived. A derivation of $\hat{w}$ and $\hat{\beta}_2$ could be achieved through an QEM iterative procedure based on the self-consistency equation. However, this would defeat the purpose as same algorithm delivers an exact maximum likelihood solution and hypotheses testing for the full model as described above. For computational simplicity, we examined an approximation where only the first-step estimators are used in the proposed score statistic. The procedure is as follows. At first step, $\hat{w}$ is taken under H0 where the cumulative baseline hazard is replaced by the Nelson Aalen estimator and $\hat{\beta}_2$ is taken as the partial likelihood estimator obtained in the corresponding time-dependent PH model. At second step, β_2 thus obtained, is used to update $\hat{w}$ using the self-consistency equation in the form of a Nelson-Aalen-Breslow estimator. It was shown by simulations that the resulting statistics is approximately χ^2 distributed with one degree of freedom. We refer to [Broët et al., 2004] for details.

8 Data analysis and properties of the QEM-based estimates

8.1 Real data examples of compound models

As another example, we use SEER data on 39393 cases of primary prostate cancer diagnosed in Greater San Francisco between 1973 and 2000. Prostate cancer specific survival was analyzed by stage of the disease (localized/regional, 35230 patients, vs. distant, 4163 patients).

Two basic models PH and PO, and two hierarchical compound models produced by compositions of PH and PO model generating functions, Γ–frailty model (20), and the PHPH cure model (16), were applied to fit the data. Stage of the disease was represented through two indicator dummy variables combined into a vector $\boldsymbol{z}$. Local/Regional stage was considered as a baseline group and the corresponding regression coefficient restricted to 0 for identifiability. Regression coefficient β for the distant stage codes for the difference in survival between the two stages expressed either as a log hazards or log odds ratio, dependent on the type of model generating function where it is used. The basic models have one predictor $\theta(\beta, z) = \exp(\beta z)$, where z=Indicator("Distant stage"). Compound models have two predictors, $\theta(\beta_\theta, z) = \exp(\beta_\theta z)$ and $\eta(\beta_\eta, z) = \exp(\beta_\eta z)$ coding two hazard ratios, long-term effect and short-term effect, respectively, in the PHPH cure

model, and odds (θ) and hazard (η) ratios in the Γ-frailty model. In the latter model, odds and hazard ratio predictors have the interpretation of the scale and shape parameter of the frailty distribution, respectively. Regression coefficients in the PH model (β_θ) and the PH submodels of the PHPH (β_θ, β_η) and Γ-frailty models (β_η) measure the disadvantage of being in the distant stage relative to local/regional stage as a relative risk. Regression coefficient in the PO model (β_θ), and the one in the PO submodel of the Γ-frailty models measure the difference from an opposite point of relative odds of survival. Since risk and odds of survival are opposites (high risk is bad, high survival is good), these coefficients are expected to be of opposite signs for in the PO and the PH model fitted to the same data.

Observed (Kaplan–Meier) and expected model–based estimates of the survival functions by group are shown in Figure 3.

Parameter estimates and confidence intervals are shown in Table 1.

Model	Parameter	Point–estimate	Confidence interval	p-Value
PH	β_θ	2.380	(2.328,2.432)	<0.001
PO	β_θ	-3.086	(-3.162,-3.011)	<0.001
PHPH	Improper PH: β_θ Proper PH: β_η	1.065 1.788	(0.923,1.207) (1.620,1.956)	<0.001 <0.001
Γ-frailty	PO: β_θ PH: β_η	-3.369 -0.179	(-3.580,-3.158) (-0.301,-0.057)	<0.001 <0.001

Table 1: Parameter estimation and hypothesis testing for prostate cancer data based on PH, PO, PHPH and Γ–frailty models. Negative β in the PO effect and positive β in the PH effect correspond to worse survival and vise versa.

Confidence intervals and hypotheses testing is based on the inverse of the observed profile information matrix.

From Figure 3 it is evident that Γ-frailty model provides the best fit to the data. The PO model is second best. Given the hierarchical structure of Γ-frailty model, its goodness of fit can be tested vs. the PO model. This is a test for $\beta_\eta = 0$ in Γ-frailty model, and it results in a significant difference $\chi^2_1 = 7.50$, $p = 0.006$. The deviance with all other models exceeds 60, and we focus on the Γ-frailty model as the best choice at the level of model complexity considered so far. We could have tried to improve on the fit by using compositions of three or more submodels, but felt that the improvement over the Γ frailty model would be irrelevant for our data. All models indicate a highly significant effect of stage ($p < 0.0001$), which is a trivial conclusion in this case.

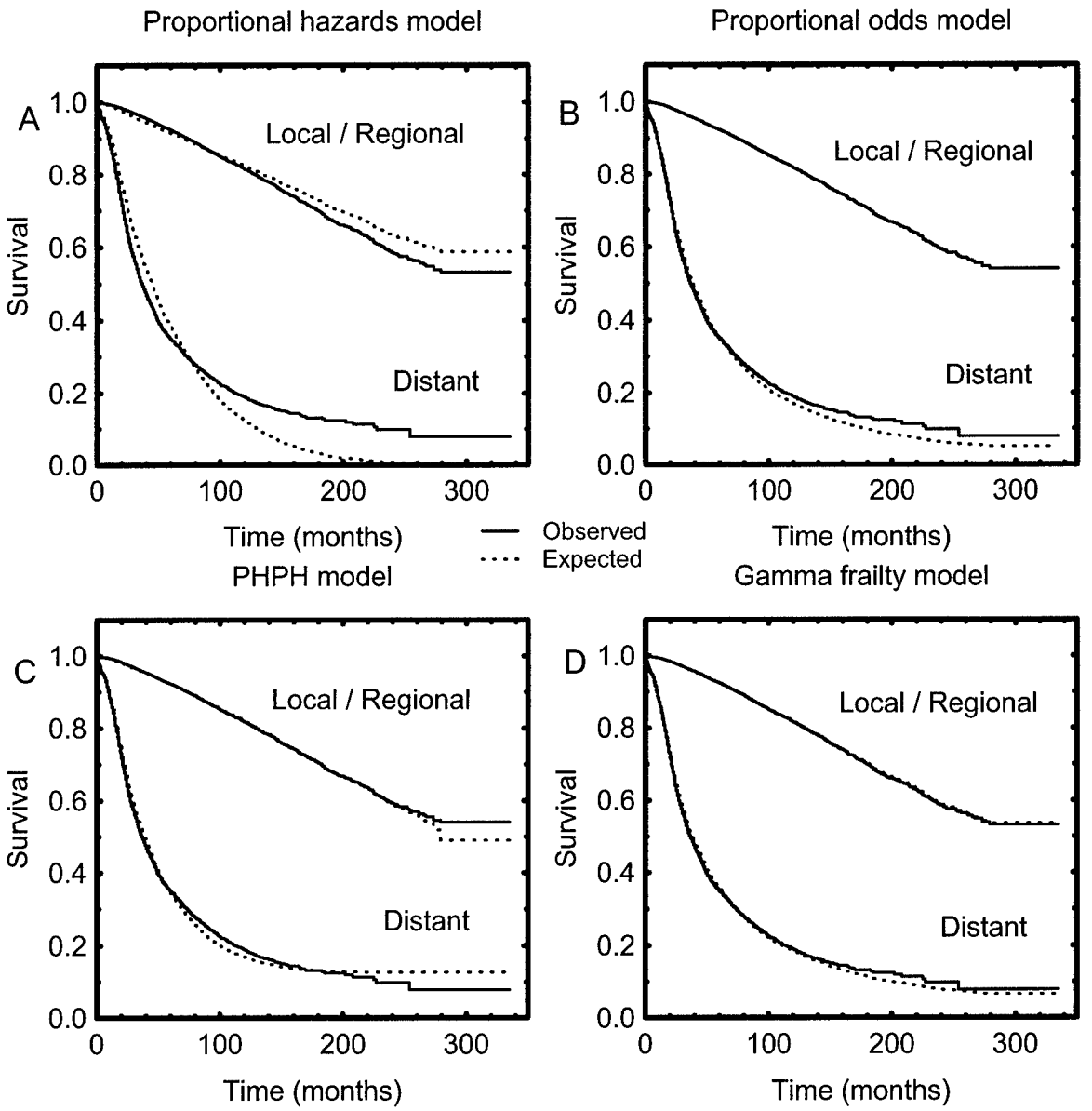


Figure 3: Prostate cancer cause-specific survival by stage. Observed (Kaplan-Meier) and expected survival curves for four models.

The validity of standard maximum likelihood theory as applied to the Γ-frailty model (20) will be studied by simulations further in this report. As the first observation, in Figure 4 we show that the form of the profile likelihood ℓ_{pr} in regression coefficients β_η (log hazards ratio) and β_θ (log odds ratio) is remarkably quadratic. In the next section we will verify by simulations that the curvature of the profile likelihood surface leads to consistent estimates of the standard errors of $\hat{\beta}$.

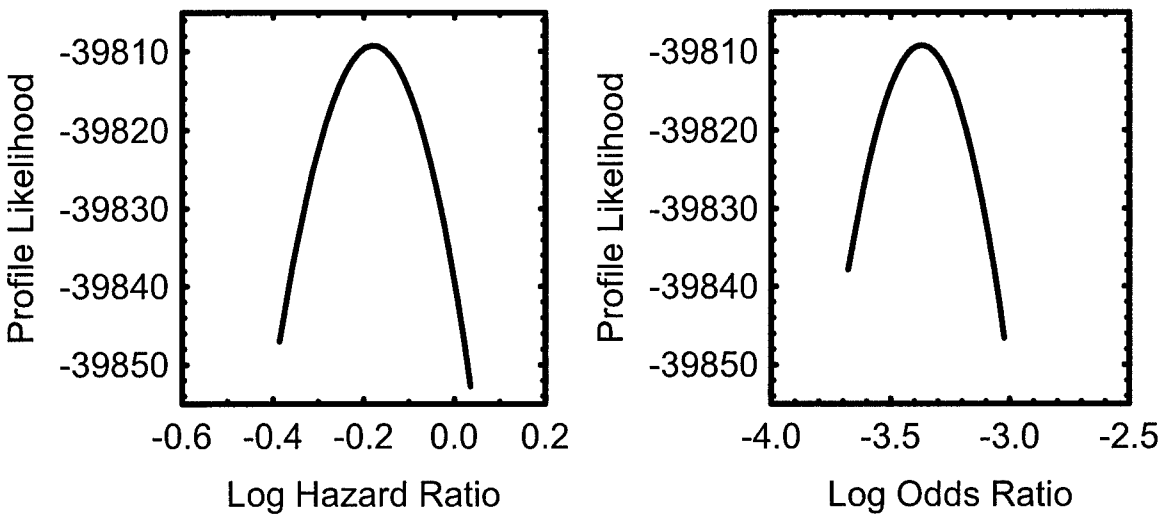


Figure 4: Profile likelihood as a function of regression coefficients sampled around the MLE point.

8.1.1 Crossing survival curves

The potential and flexibility of the PHPH model is illustrated in the following real data example of crossing survival curves [Wendland et al., 2004]. In 9 SEER registries, 8,036 females were identified who were diagnosed with Hodgkin's Disease (HD) between 1973 and 1999. Of these women, 183 (2.3%) were subsequently diagnosed with breast cancer. The use of radiation therapy in the treatment of HD resulted in an increased risk of development of breast cancer (SIR=1.90, p¡0.01). The Kaplan-Meier curves for women treated with and without radiation therapy cross at roughly 18 years after the diagnosis of HD (Figure 5). The log-rank test and proportional hazard regression model failed to detect a difference (p=0.79) in breast cancer free survival. Figure 6 demonstrates that the expected survival curves under the PH model are virtually the same for the two groups. The PHPH regression model and software developed in this project in Year 2 revealed that the use of radiation therapy had an adverse effect on long-term survival (relative risk [RR] =1.84, p=0.01), but was associated with a short-term survival benefit (RR=0.45, p=0.01). Use of the PHPH model and algorithms for hypotheses testing reported above indicates that the use of radiation therapy in the treatment of HD results in an increased long-term risk for the subsequent development of breast cancer, but confers a short-term benefit. The observed and expected survival curves using the PHPH model are in a very good agreement (Figure 7).

With the preliminary data analysis of prostate cancer both Gamma frailty model and the PHPH model provide a reasonable fit and sensitivity to the observed effects. The breast cancer example was invoked above to highlight a situation where the PHPH model is superior. While the Gamma frailty model would be a better fit than the PH, it still fails to reproduce crossing

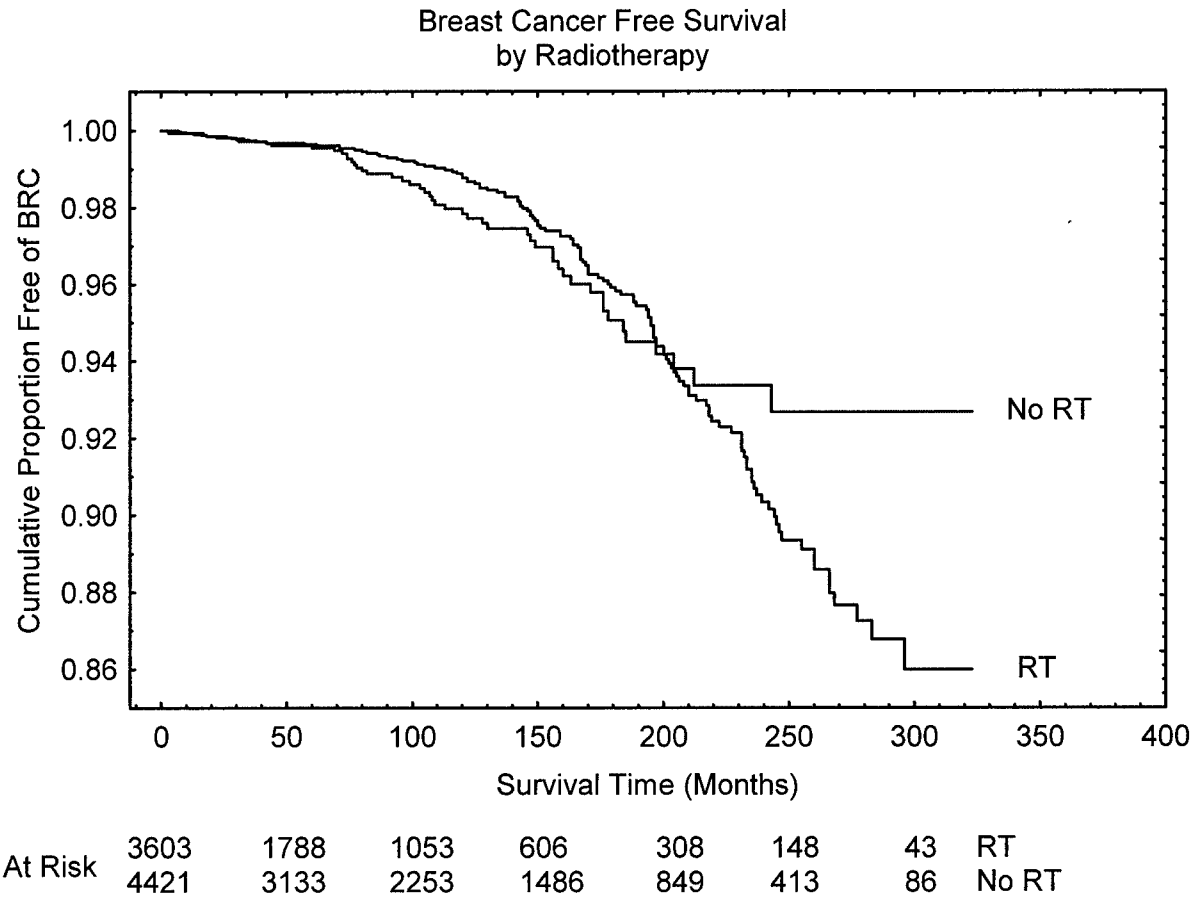


Figure 5: Kaplan-Meier curves for women treated with and without radiation therapy

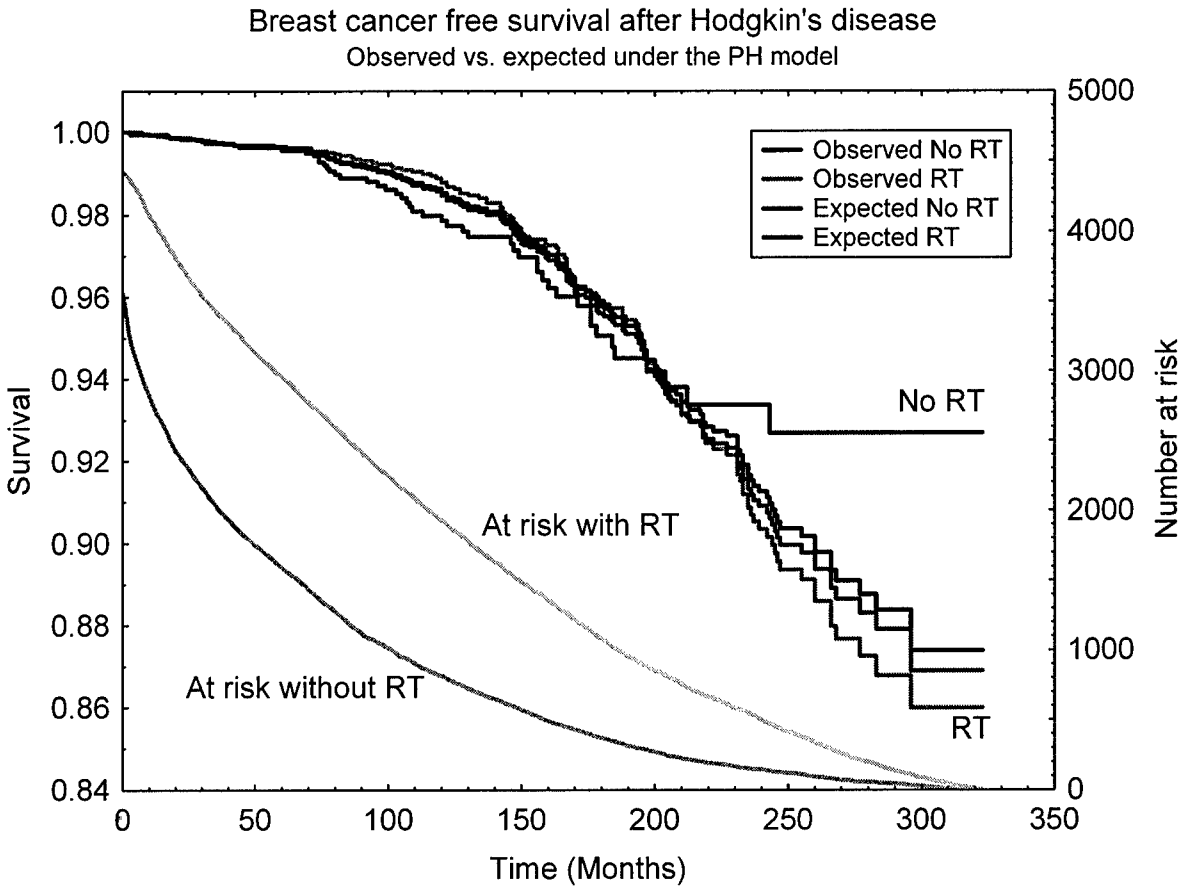


Figure 6: Observed Kaplan-Meier curves for women treated with and without radiation therapy, and their expected counterparts under the PH model.

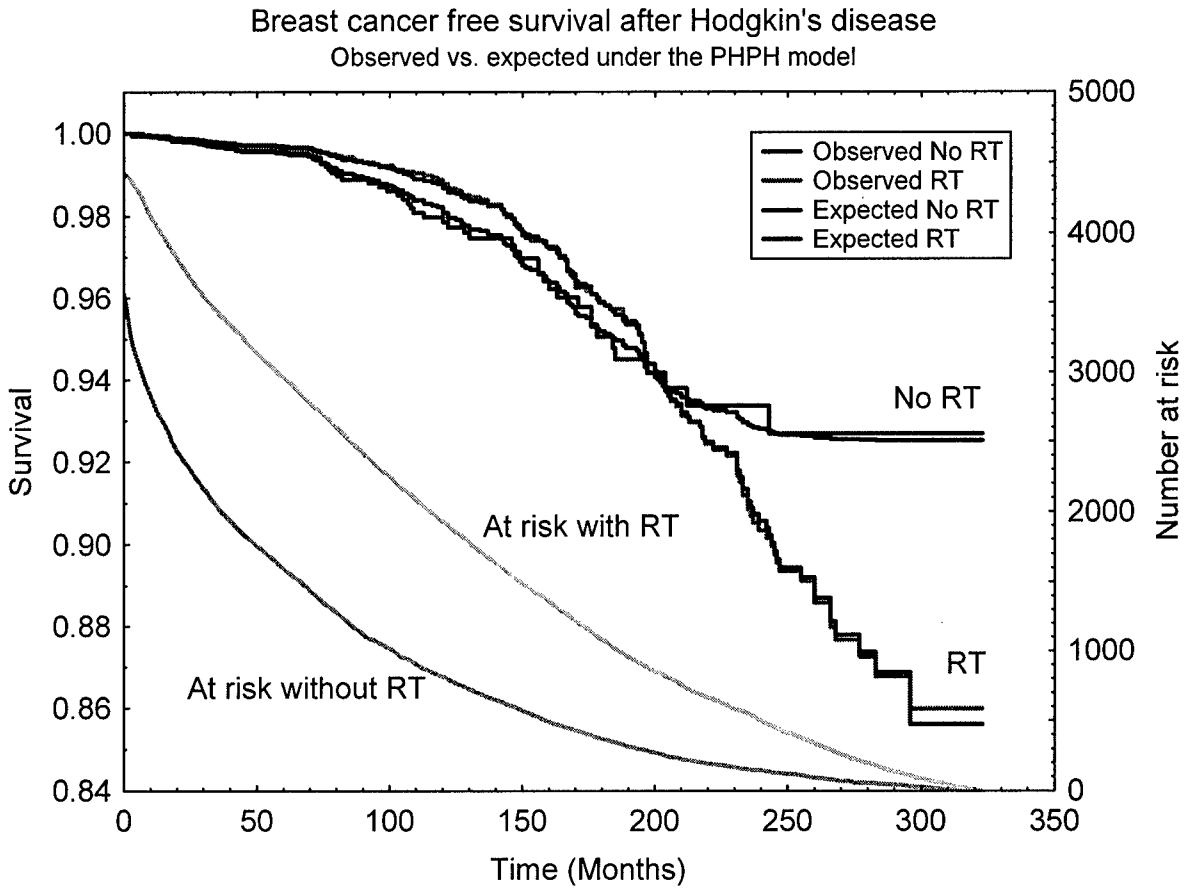


Figure 7: Observed Kaplan-Meier curves for women treated with and without radiation therapy, and their expected counterparts under the PHPH model.

curves observed in Figure 5 (not shown). We intend to keep both models in our prostate cancer analytic arsenal in case some carefully defined subsets of prostate cancer patients demonstrate similar effects.

8.2 Simulations

We begin by fitting a parametric Γ-frailty model (20) to the prostate cancer data with the baseline survival function specified as Weibull distribution. The fit (not shown) is very similar to the semiparametric version of the model, and the parameter estimates are as follows, $\beta_\theta = -3.454$, $\beta_\eta = -0.215$, and [median of F]=265.571, [shape of F]=1.491.

Each simulation experiment was replicated 1000 times. Four sets of experiments were generated with samples sizes of 100 to 1000. Shown in Figure 8 are normal probability plots for the components of $\boldsymbol{\beta} = (\beta_\theta, \beta_\eta)^{\mathrm{T}}$. As evident from the figure, small sample size may be associated with some departure from normality of MLEs, however, with a sample size larger than 300 the estimates look perfectly normal. Shown in Table 2 are the results of simulations evaluating bias and variance of the estimates. Empirical means of $\hat{\boldsymbol{\beta}}$ show good correspondence to the true parameter values used to simulate the data and are within the margin of error expected from 1000 replicates. Empirical standard errors $\mathrm{S}_n\{\hat{\beta}\}$ estimated from replicated regression coefficients are in excellent correspondence with the $\mathrm{E}_n\{\hat{\sigma}_\beta\}$, the empirical mean of the replicated I_{pr}-based estimate of standard errors. The precision of variance estimation $\mathrm{S}_n\{\hat{\sigma}_\beta\}$ improves rapidly with the sample size.

Parameter	$\mathrm{E}_n\{\hat{\beta}\}$	$\mathrm{S}_n\{\hat{\beta}\}$	$\mathrm{E}_n\{\hat{\sigma}_\beta\}$	$\mathrm{S}_n\{\hat{\sigma}_\beta\}$	Sample size
PH: β_θ	-3.168	1.394	1.451	0.415	100
PO: β_η	0.117	0.725	0.700	0.481	
PH: β_θ	-3.478	0.741	0.744	0.072	300
PO: β_η	-0.113	0.308	0.304	0.058	
PH: β_θ	-3.352	0.535	0.553	0.034	500
PO: β_η	-0.159	0.220	0.228	0.026	
PH: β_θ	-3.433	0.392	0.391	0.018	1000
PO: β_η	-0.197	0.158	0.157	0.012	

Table 2: The results of computer simulation to verify asymptotic properties of profile likelihood based MLEs.

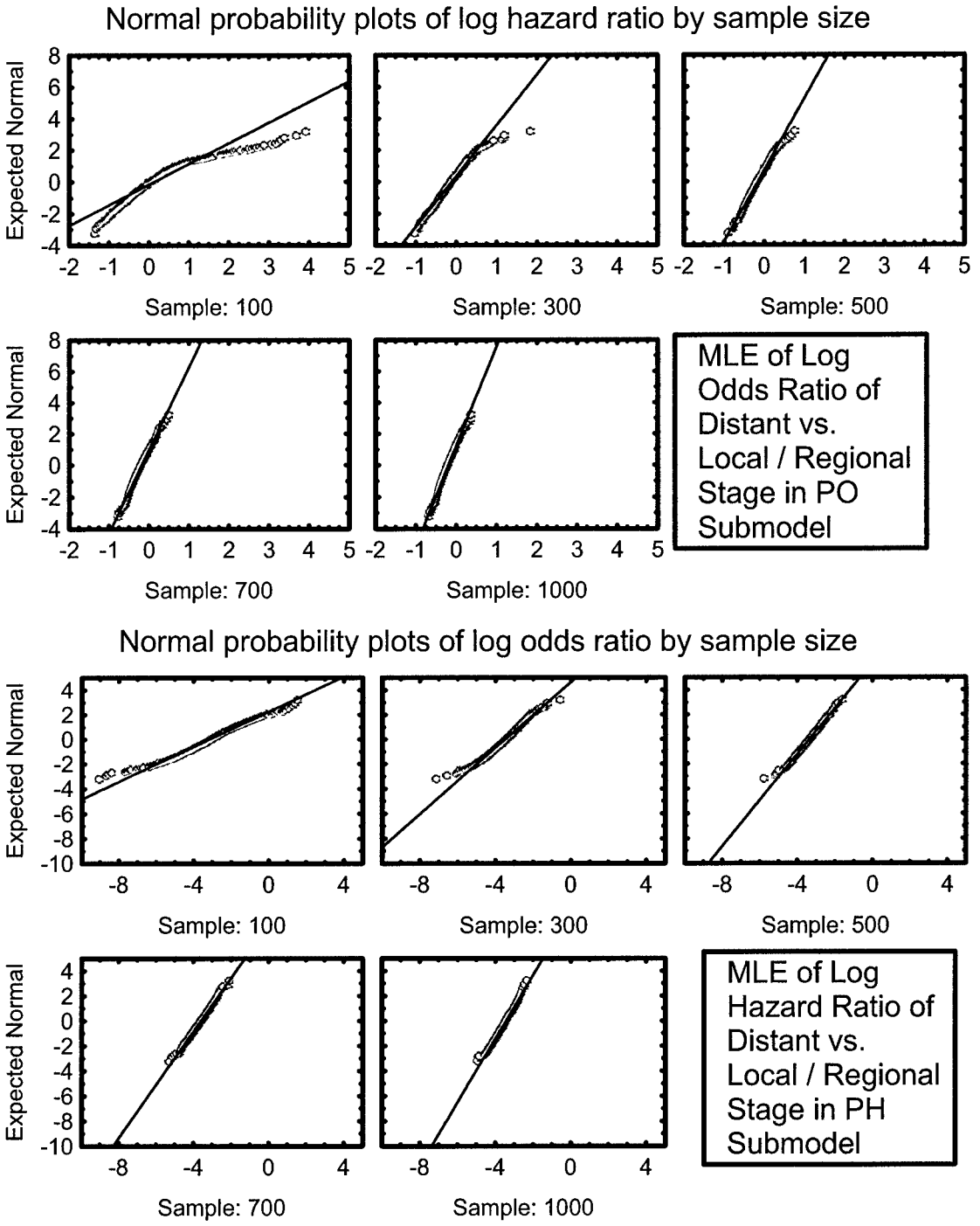


Figure 8: Normal probability plots

9 Software and data analysis

In this section we give an example of multivariate analysis of prostate cancer specific survival. A random subset of approximately 10% of all localized cases of prostate cancer diagnosed in San Francisco Bay area from 1973 to 2000 was used to test the program. The test dataset has 2751 cases of prostate cancer. The following covariates were used in the multivariate analysis.

1. Grade, "g2", 1 = Low grade (baseline), 2 = High grade;
2. Radiotherapy (any type), "Rx", 0 = No, 1 = Yes;
3. Surgery, "s2", 1 = Local or no surgery, 2 = Radical prostatectomy or En Bloc resection;
4. Race, "Race", 1 = White, 2 = Black, 3 = Other
5. Age, "Age", continuous variable, years.

The following is a fragment of the data file in the text format required by the program.

```
Time
18      Number of covariates
DxY
Race
Registry
Grade
g3
g2
Stage
Rx
Surgery
s2
s3
Agegr
Age
AgeScans
TimeScans
dxyscans
AT
ATC
0       Weight 1 = present, 0 = No weight
0   0   2000    3   1   2   2   1   1   1   0   1   0   69  69  19  0   100 3.452   4
0   0   1997    1   1   2   2   1   1   0   0   1   0   59  59  9   0   97  2.84    3
0   0   2000    1   1   2   2   1   1   0   3   2   2   69  66  16  0   100 3.404   4
0   0   1994    1   1   1   1   1   1   0   1   1   1   79  79  29  0   94  6.729   7
0   0   2000    1   1   1   1   1   1   0   0   1   0   79  77  27  0   100 3.644   4
0   0   2000    1   1   1   1   1   1   0   0   1   0   74  74  24  0   100 3.5 4
0   0   2000    1   1   2   2   1   1   0   0   1   0   64  61  11  0   100 3.062   4
0   0   2000    1   1   2   2   1   1   0   0   1   0   84  83  33  0   100 4.627   5
```

As the program is loaded, the user is brought to the input page (Figure 9). A double-click on the file name box allows us to browse to the file to be analyzed. Read data button reads the data in. At this stage, the software evaluates data for integrity, selects cases admissible to the analysis, mines the data for groups defined by categorical covariates, and calculates descriptive statistics. In the data input process the user will specify a classification of covariates into continuous or categorical and the contrasts/model used to code categorical variables. Currently full factorial and main effects are implemented. On the Covariates/Select groups page and the Curves page,

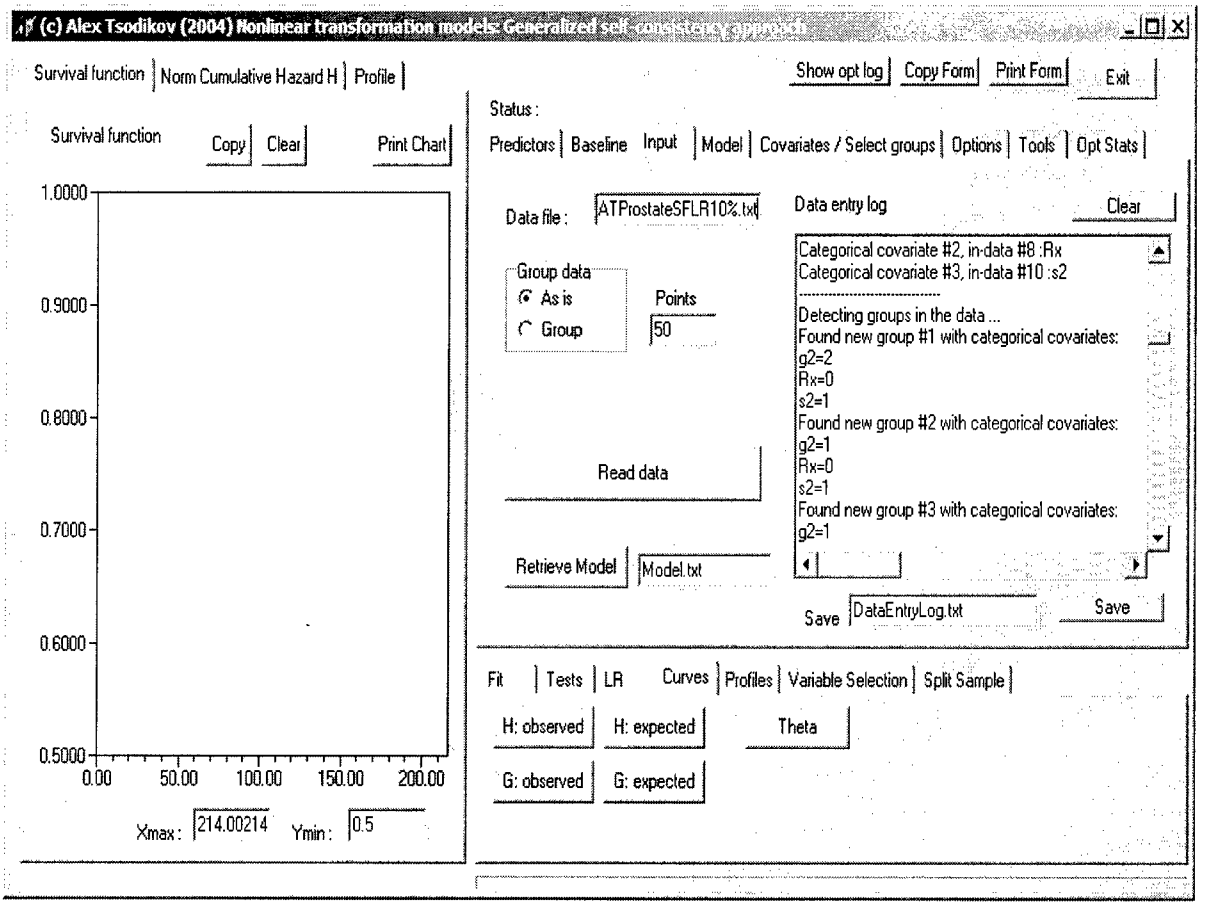


Figure 9: Data input page

the user can look at Kaplan-Meier and other descriptive curves corresponding to groups selected from the factorial strata defined by categorical variables. Figure 10 shows output for two selected groups corresponding to low grade tumors treated by surgery (No Radiotherapy) grouped by No or Local Surgery vs. Radical Surgery. The hypothesis generated by this analysis is the benefit of radical prostatectomy in localized prostate cancer vs. watchful waiting or local surgery. On the Model and Variable Selection pages (Figure 11) the user specifies the model to be used in the analysis (PHPH in the example) and variable selection procedure. The Fit page is used to specify the method (QEM algorithm in the example) and to launch model fitting. Predictors page will show final estimates and confidence intervals for the model parameters (Figure 12). Intermediate output of the model fitting procedure and variable selection is shown in Figure 13. The same figure shows the model restrictions block of the model specification. This page can be used to

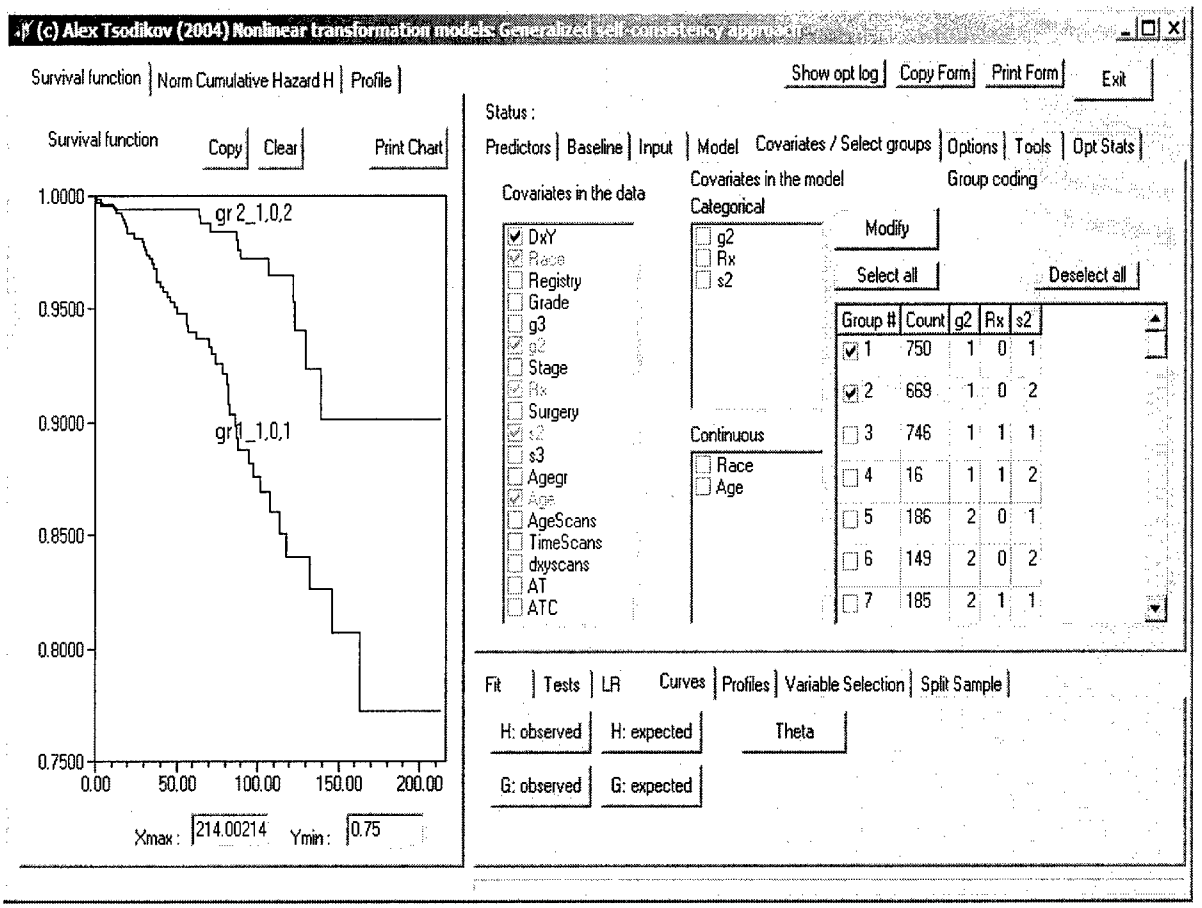


Figure 10: Descriptive slicing of the data

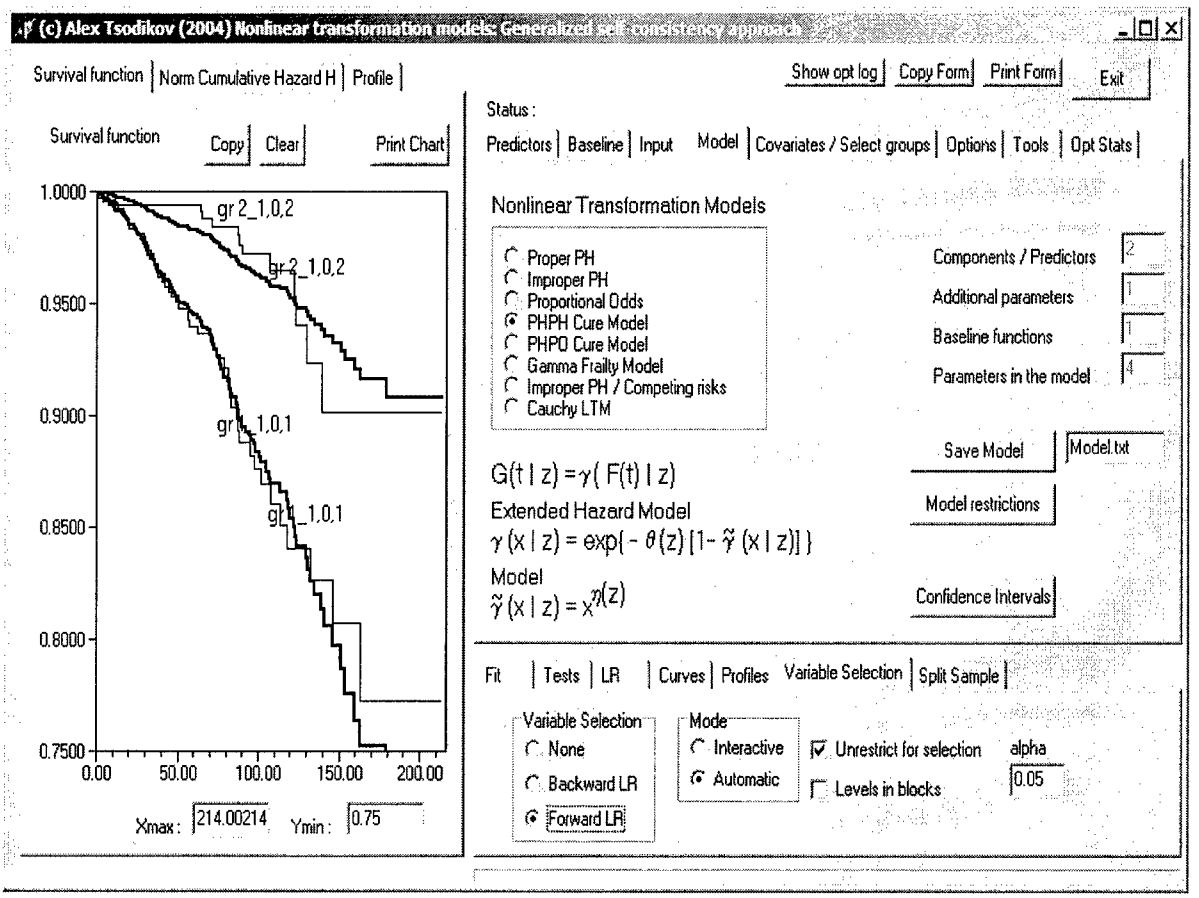
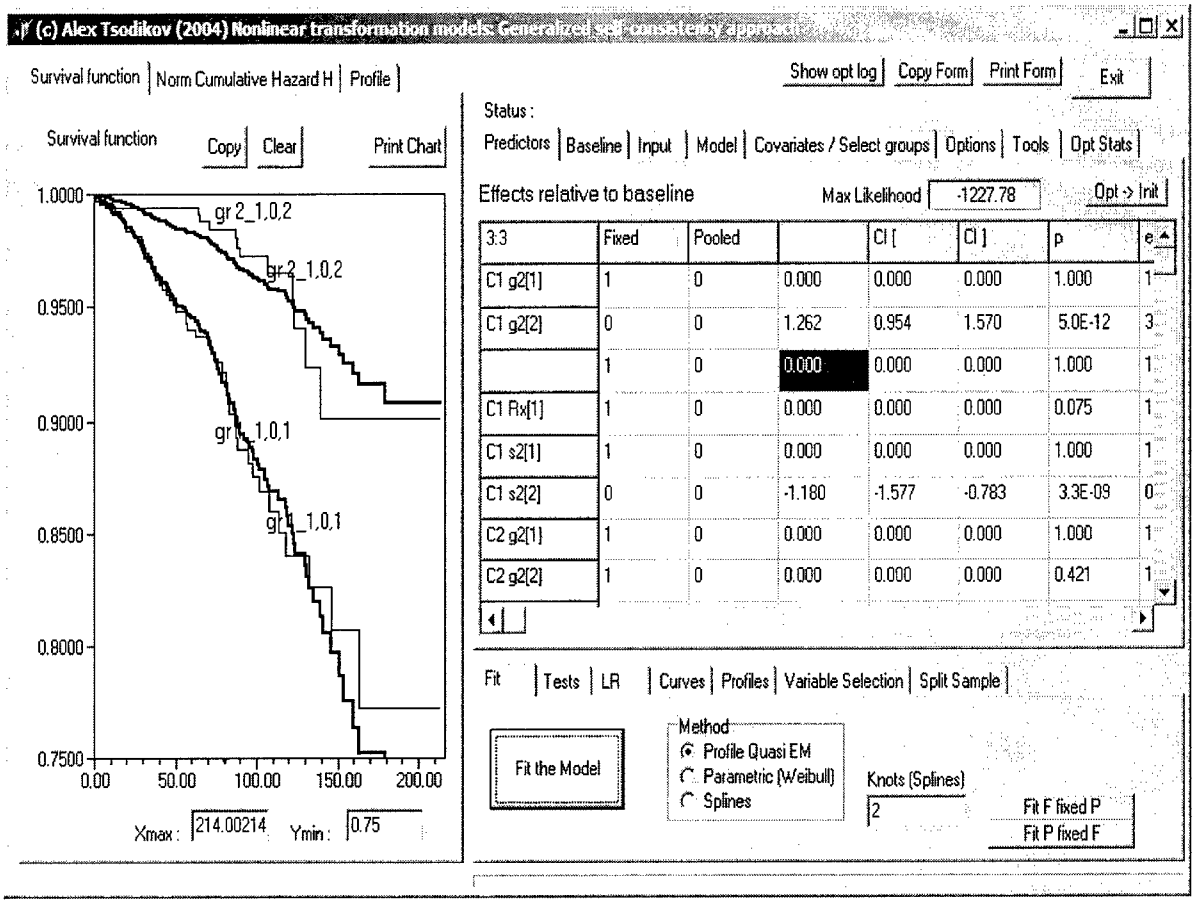


Figure 11: Choosing the model and variable selection methods. Figure in the left part of the worksheet shows observed and expected survival curves under the model.

arbitrarily fix or pool any model parameters. This functionality can be used in an automatic mode in variable selection procedures or manually, when we want to test a specific hypothesis. Model comparisons for any two hierarchical models is done by the likelihood ratio test on page LR.



3:3	Fixed	Pooled		CI [	CI]	p	e
C1 g2[1]	1	0	0.000	0.000	0.000	1.000	1
C1 g2[2]	0	0	1.262	0.954	1.570	5.0E-12	3
	1	0	0.000	0.000	0.000	1.000	1
C1 Rx[1]	1	0	0.000	0.000	0.000	0.075	1
C1 s2[1]	1	0	0.000	0.000	0.000	1.000	1
C1 s2[2]	0	0	-1.180	-1.577	-0.783	3.3E-09	0
C2 g2[1]	1	0	0.000	0.000	0.000	1.000	1
C2 g2[2]	1	0	0.000	0.000	0.000	0.421	1

Figure 12: Estimates of model parameters and confidence intervals.

Shown in table 3 the final model.

Summarizing the preliminary analysis, we observed an adverse long-term effect of high grade on survival and no short-term effect, which indicates that grade follows a PH model. Radiotherapy showed a short-term benefit, but no effect on the cure rate. Radical surgery was superior to no or local surgery in improving the chance of cure.

10 Key Research Accomplishments

Simmarizing, the key research accomplishments in Year 2 are:

1. Recurrent solution of the system of equations for variance estimation
2. A study of the meaning of the imputation operator in the QEM algorithm
3. A study of diversity of responses reproduced by compound Nonlinear Transformation Models

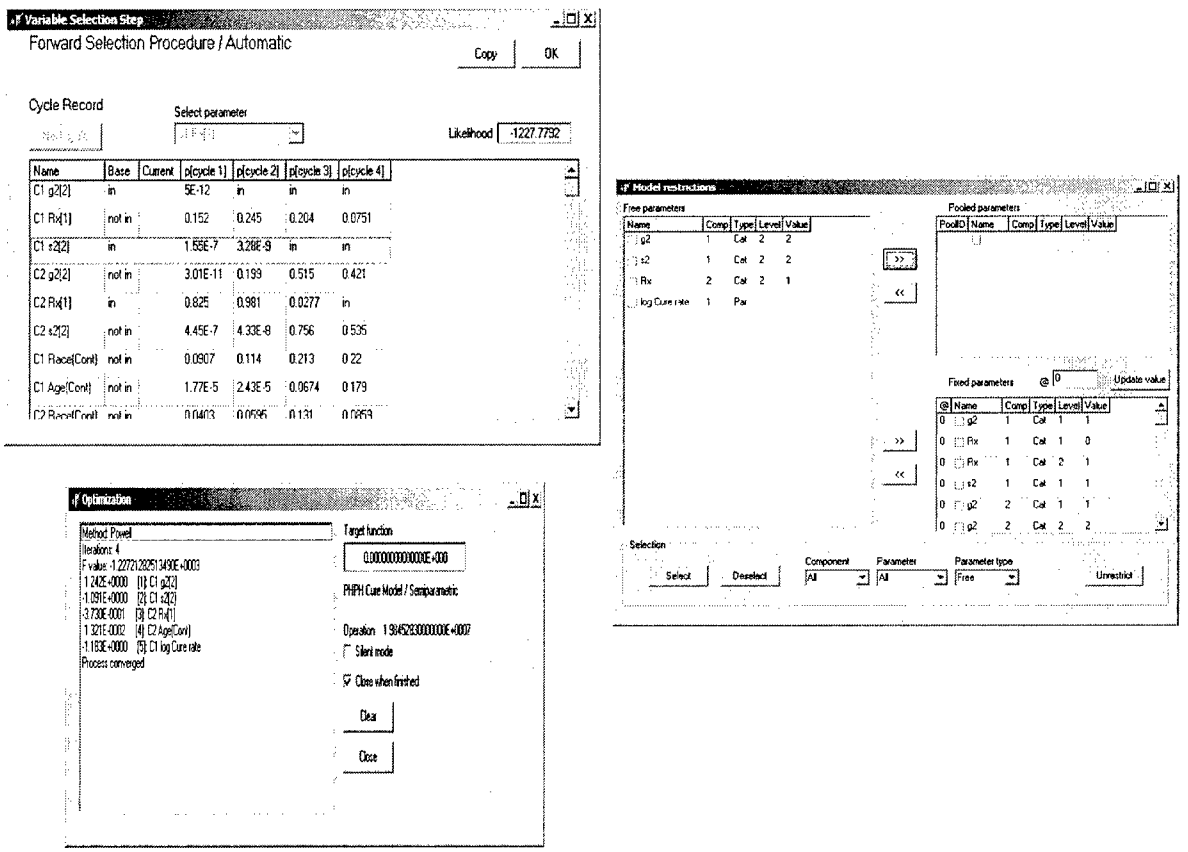


Figure 13: Interim output of the model fitting and variable selection algorithms (left). Model restrictions window (right).

Effect	Parameter	Point–estimate	Confidence interval	p-Value
LT	High vs. Low grade	1.262	(0.954,1.570)	<0.001
LT	No or Local vs. Radical Surgery	-1.180	(-1.577,-0.783)	<0.001
LT	Baseline Log cure rate	-1.162	(-1.528,-0.796)	<0.001
ST	Radiotherapy Yes vs. No	-0.423	(-0.801,-0.045)	0.028

Table 3: Parameter estimates in the final PHPH model for the test dataset. "LT" = Long-term effect, "ST" = Short-term effect. Negative βs correspond to lower risk.

4. A score test for long- and short-term effects

5. Computer implementation of point and interval estimation, hypothesis testing and variable selection based on multivariate semiparametric nonlinear transformation models

6. Study of properties of estimators by simulation

7. Multivariate regression analysis of localized prostate cancer specific survival based on population registry data.

11 Reportable Outcomes

11.1 Manuscripts

1. Tsodikov, A. (2004) Generalized self-consistency methods for cure models, In "Recent developments in censored data analysis" INSERM, Paris, 2004.

2. Broët, P., Tsodikov, A., De Rycke, Y., Moreau, T. (2004) Two-sample statistics for testing the equality of survival functions against improper semi-parametric accelerated failure time alternatives: An application to the analysis of a breast cancer clinical trial, Lifetime Data Analysis, Vol. 10, 103-120.

3. Wendland, M.M., Tsodikov, A., Glenn, M.J., Gaffney, D.K. (2004) Time interval to the development of breast cancer following treatment for Hodgkin's Disease, Cancer, Vol. 101, 1275-1282.

11.2 Presentations

1. Tsodikov, A. (2004) Cure Models (invited), Workshop of the French National Institutes of Health (INSERM).

2. Tsodikov, A. (2004) Modeling and estimation of cancer incidence and mortality under variable dissemination of screening with application to prostate cancer (invited), International Biometric Conference, Cairns, Australia July 2004.

3. Tsodikov, A. (2004) Population impact of PSA testing. The Tenth Annual Cancer Research Symposium October 20-21, UCD Cancer Center.

12 Conclusions

In Year 2 we have completed methodology and software development for point and interval estimation and variable selection for compound Nonlinear Transformation Models. We have built a number of candidate compound models for prostate cancer and verified their properties analytically and by simulations. Finally, we used the new software and methodology to apply these models to a number of real and simulated test data sets.

In the last Year 3 of the project we will focus on large scale analysis of real prostate data on cancer specific survival and biochemical recurrence. We are planning to evaluate utility of tree-based models or alternative strategies of best model selection and analysis of treatment-covariate interactions. A high-performance computer workstation will be purchased when computer intensive methodology has been incorporated into the software to deal with computational challenges of best model selection and analysis of interactions. This analysis has the goal of identifying subsets of patients with indication for particular treatment.

13 References

References

O.O. Aalen. Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability*, 2:951–972, 1992.

P. Broët, A. Tsodikov, Y. De Rycke, and T. Moreau. Two-sample statistics for testing the equality of survival functions against improper semi-parametric accelerated failure time alternatives: An application to the analysis of a breast cancer clinical trial. *Lifetime Data Analysis*, 10:103–120, 2004.

S.A. Murphy and A.W. van der Vaart. On profile likelihood. *Journal of the American Statistical Association*, 95:449–485, 2000.

G.G. Nielsen, R.D. Gill, P.K. Andersen, and T.I. Sorensen. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19:25–43, 1992.

W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipies in Pascal. The Art of Scientific Computing.* Cambridge University Press, New York, NY, 1994.

A. Tsodikov, J.G. Ibrahim, and A.Y. Yakovlev. Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association*, 98:1063–1078, 2003.

A. Tsodikov. Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society, Series B*, 65:759–774, 2003.

J.W. Vaupel, K.G. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439–454, 1979.

M.M. Wendland, A. Tsodikov, M.J. Glenn, and D.K. Gaffney. Time interval to the development of breast cancer following treatment for hodgkin's disease. *Cancer*, 101:1275–1282, 2004.

14 Appendix

List of papers presented in the appendix

1. Tsodikov, A. (2004) Generalized self-consistency methods for cure models, In "Recent developments in censored data analysis" INSERM, Paris, 2004.

2. Broët, P., Tsodikov, A., De Rycke, Y., Moreau, T. (2004) Two-sample statistics for testing the equality of survival functions against improper semi-parametric accelerated failure time alternatives: An application to the analysis of a breast cancer clinical trial, Lifetime Data Analysis, Vol. 10, 103-120.

3. Wendland, M.M., Tsodikov, A., Glenn, M.J., Gaffney, D.K. (2004) Time interval to the development of breast cancer following treatment for Hodgkin's Disease, Cancer, Vol. 101, 1275-1282.

Ateliers de formation
N°154

Inserm

Institut national
de la santé et de la recherche médicale

Développements récents en analyse de données censurées

Recent developments in censored data analysis

1er-2 juillet 2004

July 1st – 2nd, 2004

La Londe-Les-Maures

Generalized Self-Consistency Methods for Cure Models

A. Tsodikov

April 29, 2004

SUMMARY

1 Introduction

A large class of semiparametric survival models can be represented by the survival function $G(t \mid \boldsymbol{z})$ given covariates $\boldsymbol{z}$ treated as a function of an unspecified baseline survival function F (or the corresponding cumulative hazard function $H = -\log F$), and a vector of regression coefficients $\boldsymbol{\beta}$. With such semiparametric models, we present a unified approach for model building and construction of numerically efficient algorithms for maximum likelihood inference. The approach is based on a generalization of the idea of self-consistency and is motivated by frailties and the EM algorithm. Composition technique is developed for building hierarchical model families compatible with the algorithms. An algorithm is provided to obtain the exact profile information matrix for the parametric part of the model. The approach is illustrated using cure models and real data.

2 Frailty models

2.1 PH mixture model

For a survival function $G(t \mid \boldsymbol{\beta}, \boldsymbol{z})$, where $\boldsymbol{\beta}$ are regression coefficients, and $\boldsymbol{z}$ are covariates, consider a PH mixture model

$$G(t \mid \boldsymbol{\beta}, \boldsymbol{z}) = \mathrm{E}\left\{ F(t)^{U(\boldsymbol{\beta},\boldsymbol{z})} \,\middle|\, \boldsymbol{z} \right\}, \tag{1}$$

where F is the baseline survival function, and $U(\boldsymbol{\beta}, \boldsymbol{z})$ is a nonnegative random variable whose distribution depends on covariates and regression coefficients. The family (1) generates cure models if U has a point-mass at zero, which corresponds to a fraction of immune individuals $\Pr\{U = 0\} > 0$ showing zero risk. Binary variable U has been a popular choice (Farewell, 1982; Kuk and Chen, 1992; Peng and Dear,

2000; Sy and Taylor, 2000). Perhaps a more attractive class of cure models stems from the compound Poisson structure for U Tsodikov et al. (2003), which we will consider later as an example.

We can make the following important observations about the class of PH mixture models (1):

- The survival function (1) is built by composition

$$G(t \mid \boldsymbol{\beta}, \boldsymbol{z}) = (\gamma \circ F)(t \mid \boldsymbol{\beta}, \boldsymbol{z}), \tag{2}$$

  where $\gamma(x \mid \boldsymbol{\beta}, \boldsymbol{z})$ is the probability generating function of U.

- The moment generating function $\gamma(x \mid \cdot)$ is a distribution function in x with the support on $[0, 1]$. If the distribution of U is specified parametrically, γ is a parametric regression model on $[0, 1]$.

- The fact that the range and the support of γ are the same allows one to build compositions of an arbitrary number of γs. As we verify in the sequel, the class of PH mixture models is closed with respect to such compositions.

These observations are used to generalize the PH mixture family into the Nonlinear Transformation Models (NTM) family (Section 4). In doing so, we make use of the following key property of the PH mixture model

Proposition 2.1 *(Tsodikov, 2003)*
Suppose, we have an observation $(t, \boldsymbol{z}, c)$ sampled from the PH mixture model under independent censoring, where t is an observed survival time and c is a censoring indicator ($c = 0$ if t is a censored survival time, and $c = 1$ if t is a failure). Then, under the PH mixture model (1),

- *the conditional expectation of U, given the observed event $(t, \boldsymbol{z}, c)$ is given by*

$$\mathrm{E}\{U(\cdot) \mid t, \cdot, c\} = (\Theta \circ F)(t \mid \cdot, c) = \Theta[F(t) \mid \cdot, c],$$

  where the function Θ is given by

$$\Theta[x \mid \cdot, c] = c + x \frac{\gamma^{(c+1)}(x \mid \cdot)}{\gamma^{(c)}(x \mid \cdot)}, \tag{3}$$

  where $\gamma^{(c)}(x \mid \cdot) = \partial^c \gamma(x \mid \cdot)/\partial x^c$, $c = 0, 1, \ldots$, $\gamma^{(0)}(x \mid \cdot) = \gamma(x \mid \cdot)$.

- *The function* $\Theta[x \mid \cdot, c]$ *is nondecreasing in* x *for any* $c = 0, 1$.

2.2 Example

We are extending the example of Tsodikov (2002). Consider the frailty variable U constructed as

$$U = \eta(\boldsymbol{z})V,$$

where $V \sim \text{Poisson}(\theta(\boldsymbol{z}))$. It is straightforward to verify that this leads us to the following model

$$\bar{G}(t|\boldsymbol{\beta}, \boldsymbol{z}) = \exp\left[-\theta(\boldsymbol{\beta}, \boldsymbol{z})\left\{1 - F(t)^{\eta(\boldsymbol{\beta},\boldsymbol{z})}\right\}\right]. \tag{4}$$

The model (4) was proposed by Broët et al. (2001) in the context of two sample score tests for long- and short-term covariate effects, see also (Tsodikov, 2002; Tsodikov et al., 2003) for applications and more discussion. We use breast cancer data from the SEER program (http://seer.cancer.gov/) to illustrate the frailty underpinnings of model (4).

The nondecreasing character of the function Θ in (3) is quite natural. The longer the subject stays event–free, the lower the subject's posterior risk, represented by Θ.

2.3 Composition with PH mixture models

The idea to use compounding to build particular extended families of frailty models is not new. For example, Aalen (1992) used a compound Poisson distribution to extend a class of frailty models by Hougaard (1984).

Consider the following general compounding techniques for the PH mixture model. If ν is a nonnegative discrete random variable with the moment generating function $\gamma_\theta(x) = \mathrm{E}\{x^\nu\}$, and ξ_k are i.i.d. copies of another nonnegative random variable (independent of ν) with the the moment generating function $\gamma_\eta(x) = \mathrm{E}\{x^\xi\}$, and U is a compound random variable given by

$$U = \sum_{k=1}^{\nu} \xi_k, \tag{5}$$

then by the composition property of Laplace transform,

$$\gamma(x) = \mathrm{E}\left\{x^U\right\} = (\gamma_\theta \circ \gamma_\eta)(x). \tag{6}$$

A large variety of semiparametric mixture models can be derived from (6). When $\gamma_\theta(x)$ corresponds to a continuous random variable, the composition $\gamma_\theta \circ \gamma_\eta$ still leads to a PH mixture model.

Proposition 2.2 *Composition for mixture models.*
Let γ_θ and γ_η be some two mixture models $\gamma_\theta(x|\cdot) = E(x^\nu \mid \cdot)$, $\gamma_\eta(x|\cdot) = E(x^\xi \mid \cdot)$, where ν and ξ are some independent nonnegative random variables. Let $\gamma = \gamma_\theta \circ \gamma_\eta$ be the compound model. Then γ is also a mixture model, meaning that there exists a nonnegative random variable U such that $\gamma(x|\cdot) = E(x^U \mid \cdot)$.

Coming back to our example, we see that the model (4) is composed of γ_θ representing a probability generating function of Poisson distribution, and γ_η representing a non-random real number η.

3 Profile likelihood approach

The problem of Nonparametric Maximum Likelihood Estimation (NPMLE) with the semiparametric model is to find estimates of regression coefficients $\boldsymbol{\beta}$, and an NPMLE estimate of H such that they deliver the maximum of a suitably defined likelihood function $\ell = \ell(\boldsymbol{\beta}, H)$. We use a profile likelihood approach to maximize ℓ. The profile likelihood is defined as a supremum of the full likelihood taken over the nonparametric part of the model

$$\ell_{pr}(\boldsymbol{\beta}) = \max_H \ell(\boldsymbol{\beta}, H). \tag{7}$$

Assuming that we are able to find the global maximum of ℓ with respect to H, given $\boldsymbol{\beta}$, we may write the profile likelihood as an implicit function of $\boldsymbol{\beta}$

$$\ell_{pr}(\boldsymbol{\beta}) = \ell\left\{\boldsymbol{\beta}, H(\boldsymbol{\beta})\right\}, \tag{8}$$

where $H(\boldsymbol{\beta})$ is the solution of a self-consistency equation. Our algorithms will be designed following a straightforward nested procedure:

- Maximize $\ell_{pr}(\boldsymbol{\beta})$ by a conventional nonlinear programming method (for example, a directions set method).

- For any $\boldsymbol{\beta}$ as demanded in the above maximization procedure, solve the self-consistency equation.

4 Nonlinear Transformation Models

We considered semiparametric survival models of the form

$$G(t \mid \cdot) = \mathrm{E}\left\{ F(t)^{U(\cdot)} \mid \cdot \right\} = (\gamma \circ F)(t \mid \cdot),$$

where γ is a probability generating function of a nonnegative random variable U. We also noticed that $\gamma(x \mid \cdot)$ is a distribution function in $x \in [0,1]$ with the range contained in the same interval of $[0,1]$. This brings us to the following natural generalization of the PH mixture family of models.

Definition 4.1 *Let $\gamma(x \mid \boldsymbol{\beta}, \boldsymbol{z})$ be a parametrically specified distribution function with the x-domain of $[0,1]$. Let $F(t)$ be a nonparametrically specified baseline survival function. A semiparametric regression survival model is called a Nonlinear Transformation Model if its survival function can be represented in the form*

$$G(t \mid \boldsymbol{\beta}, \boldsymbol{z}) = \gamma\left\{F(t) \mid \boldsymbol{\beta}, \boldsymbol{z}\right\} = (\gamma \circ F)(t \mid \boldsymbol{\beta}, \boldsymbol{z}). \tag{9}$$

Functions γ will be called NTM-generating functions.

The class (9) was introduced in (Tsodikov, 2003), where universal estimation algorithms for the NTM class were developed. The key requirement that ensures monotonicity and convergence of the estimation algorithms of Section 5 is that of nondecreasing Θ, where Θ is defined in (3). Now that we no longer use the concept of frailty in the definition of NTM, $\Theta(F \mid \cdot, c)$ becomes a surrogate of the posterior risk such that its basic property of nondecreasing $\Theta(x \mid \cdot, c)$ is preserved.

4.1 Composition

If γ_θ and γ_η are two different NT models with predictors θ, and η, respectively, then

$$\gamma(x|\cdot) = (\gamma_\theta \circ \gamma_\eta)(x|\cdot) \tag{10}$$

is a new semiparametric model with two predictors θ and η. If $\gamma_\theta(x|\cdot) \equiv x$ for some value of θ (usually for $\theta = 1$), then the model (10) includes models γ_θ and γ_η as nested special cases. The fact that NTM–generating functions $\gamma(x \mid \cdot)$ are all defined on $x \in [0, 1]$ and have the range in the same interval allows us to compose as complex a hierarchical model as needed. Moreover, operation of composition preserves the key property of nondecreasing Θ observed in PH mixture model

Proposition 4.1 *Composition.*
Let γ_θ and γ_η be some two NTM–generating functions, each satisfying the assumption of nondecreasing Θ, where Θ is given by (3), and let $\gamma = \gamma_\theta \circ \gamma_\eta$ be the compound function (compositions are taken with respect to x). Let Θ_a be the Θ–function (3) corresponding to γ_a, $a = \theta, \eta$, and to the compound function γ, if a is blank. Then (A)

$$\Theta(x \mid \cdot, c) = \Theta_\eta(x \mid \cdot, 0)\left\{(\Theta_\theta \circ \gamma_\eta)(x \mid \cdot, c) - c\right\} + c\Theta_\eta(x \mid \cdot, c), \tag{11}$$

where $c = 0, 1$ and $(\Theta \circ \gamma)(x \mid \cdot, c)$ is understood as $\Theta\{\gamma(x \mid \cdot)|\cdot, c\}$; and
(B) The function Θ (3) derived from the compound NTM–generating function γ is nondecreasing in x as required for monotonicity and convergence of the estimation algorithms (See Section 5).

5 Estimation algorithm

Let t_i, $i = 1, \ldots, n$ be a set of times, arranged in increasing order, $t_{n+1} := \infty$. Associated with each t_i is a set of subjects $\mathcal{D}_i$ with covariates $\boldsymbol{z}_{ij}$, $j \in \mathcal{D}_i$ who fail at t_i, and a similar set of subjects $\mathcal{C}_i$ with covariates $\boldsymbol{z}_{ij}$, $j \in \mathcal{C}_i$ who are censored at t_i. The observed event $\mathcal{E}_{ij}$ for the subject ij is a triple $(t_i, \boldsymbol{z}_{ij}, c_{ij})$, where c is a censoring indicator, $c = 1$ if failure, $c = 0$ if right censored. For any function

$A(t)$, let $A_i = A(t_i)$, $\Delta A_i = |A(t_i) - A(t_i - 0)|$. A step-wise function H can be characterized by two vectors $\Delta \boldsymbol{H} = (\Delta H_1, \ldots, \Delta H_n)^{\mathrm{T}}$ and $\boldsymbol{t} = (t_1, \ldots, t_n)^{\mathrm{T}}$.

The following method (QEM) is used to obtain the profile likelihood and solve (7):

$$\Delta H_m^{(k+1)} = \frac{D_m}{\sum_{ij \in \mathcal{R}_m} \Theta(F_i^{(k)} \,|\, \boldsymbol{\beta}_{ij}, \boldsymbol{z}_{ij}, c_{ij})}, \tag{12}$$

where $\{F^{(k)}\}$ and $\{H^{(k)}\}$ are sequences of functions generated by the self-consistency equation (12).

It can be shown that if Θ is nondecreasing, each update of H using the self-consistency equation (12) strictly improves the likelihood, given $\boldsymbol{\beta}$. This guarantees convergence of the sequence of likelihood values $\ell\left\{\boldsymbol{\beta}, H^{(k)}\right\}$ to the profile likelihood of $\boldsymbol{\beta}$, and of the sequence $\{H^{(k)}\}$ to H^*, the fixed point of (12), under fairly general conditions.

Under a PH mixture model, the procedure (12) is an EM algorithm based on imputation of the missing predictor U in the Nelson–Aalen–Breslow estimator by its conditional expectation, given observed data, represented by $\Theta(F \,|\, \boldsymbol{\beta}, \boldsymbol{z}, c)$. Under an NT model, the procedure works as a Quasi-EM algorithm without the missing-data interpretation.

6 Profile information matrix

As the number of parameters of a semiparametric model is potentially unlimited, obtaining the inverse of the full information matrix can be computationally prohibitive. Therefore, we use the profile information matrix

$$\boldsymbol{I}^{\mathrm{P}}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} + \left(\frac{\partial \boldsymbol{H}^*}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{H}\boldsymbol{H}} \frac{\partial \boldsymbol{H}^*}{\partial \boldsymbol{\beta}} + \left(\frac{\partial \boldsymbol{H}^*}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}} \boldsymbol{I}_{\boldsymbol{H}\boldsymbol{\beta}} + \boldsymbol{I}^{\mathrm{T}}_{\boldsymbol{H}\boldsymbol{\beta}} \frac{\partial \boldsymbol{H}^*}{\partial \boldsymbol{\beta}}, \tag{13}$$

where

$$\boldsymbol{I}_{\boldsymbol{a}\boldsymbol{b}} = - \left. \frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{H})}{\partial \boldsymbol{a} \partial \boldsymbol{b}^{\mathrm{T}}} \right|_{(\boldsymbol{\beta}, \boldsymbol{H}^*(\boldsymbol{\beta}))}$$

for any two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, and H^* is the fixed point of the self-consistency equation.

Notice that $\boldsymbol{I}^{\mathrm{P}}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ has dimension $d \times d$, with $d = \dim(\boldsymbol{\beta})$, therefore only a small matrix needs to be inverted in order to get an estimator of the covariance matrix of

regression coefficients.

The downside of (13) is that since $\boldsymbol{H}^*(\boldsymbol{\beta})$ is defined implicitly, so is the potentially large Jacobian matrix $\partial \boldsymbol{H}^*/\partial \boldsymbol{\beta}$. Therefore, the Jacobian is generally unavailable in a closed form. In the NTM case the problem reduces to solving a system of linear equations $(\boldsymbol{D}+\boldsymbol{R})\boldsymbol{x}=\boldsymbol{b}$, where x represents a column–vector of the Jacobian, $\boldsymbol{D}$ is an $n \times n$ diagonal matrix with diagonal elements $d_i \neq 0$, $i=1,\ldots,n$, $\boldsymbol{R}=(R_{kl})$ is a $n \times n$ matrix, $R_{kl}=\sum_{i=\max\{k,l\}}^{n} a_i$, a_i, $i=1,\ldots n$ are real numbers, and $\boldsymbol{b}$ be an n-dimensional vector.

Proposition 6.1 gives the main result used to efficiently calculate $\partial \boldsymbol{H}^*/\partial \boldsymbol{\beta}$.

Proposition 6.1 *Define the functions $\varphi_k : \mathbb{R} \to \mathbb{R}$, $k=1,\ldots,n$ recursively in the following way,*

$$\begin{aligned}
\varphi_n(y) &= \frac{b_n}{d_n} - \frac{a_n}{d_n} y, \\
\varphi_k(y) &= \frac{1}{d_k}\left(b_k - \sum_{i=k}^{n} a_i y + \sum_{l=k+1}^{n} \sum_{i=k}^{l-1} a_i \varphi_l(y) \right), \quad k=n-1,\ldots,1.
\end{aligned}$$

Let $\tilde{\varphi} : \mathbb{R} \to \mathbb{R}$ be the function given by $\tilde{\varphi}(y)=\sum_{k=1}^{n} \varphi_k(y)$ and let

$$\tilde{y} = \frac{\tilde{\varphi}(0)}{1+\tilde{\varphi}(0)-\tilde{\varphi}(1)}.$$

The solution to the system of equations $(\boldsymbol{D}+\boldsymbol{R})\boldsymbol{x}=\boldsymbol{b}$ is the n-dimensional vector $\boldsymbol{x}=(\varphi_1(\tilde{y}),\ldots,\varphi_n(\tilde{y}))^T$.

Acknowledgement

This research is supported by National Cancer Institute grant U01 CA97414, and Department of Defence grant DAMD17-03-1-0034.

References

Aalen, O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability 2*, 951–972.

Cheng, S., L. Wei, and Z. Ying (1995). Analysis of transformation models with censored data. *Biometrika 82*, 835–845.

Cheng, S., L. Wei, and Z. Ying (1997). Predicting survival probabilities with semiparametric transformation models. *Journal of the American Statistical Association 92*(437), 227–235.

Clayton, D. and J. Cuzick (1985a). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society, Series A 148*, 82–117.

Clayton, D. and J. Cuzick (1985b). The semiparametric pareto model for regression analysis of survival times. *Bull. Internat. Statist. Inst. 51*, 1–18.

Dabrowska, D. and K. Doksum (1988). Estimation and testing in a two-sample generalized odds-rate model. *Journal of the Americal Statistical Association 83*, 744–749.

Doksum, K. and M. Gasko (1990). On a correspondence between models in binary regression analysis and in survival analysis. *International Statistical Review 58*, 243–252.

Farewell, V. (1982). The use of mixture models for the analysis of survival data with long–term survivors. *Biometrics 38*, 1041–1046.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* New York: John Wiley and Sons.

Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika 71*(1), 75–83.

Klein, J. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics 48*, 795–806.

Kuk, A. and C.-H. Chen (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, 531–541.

Murphy, S., A. Rossini, and A. van der Vaart (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association 92*(439), 968–976.

Murphy, S. and A. van der Vaart (1997). Semiparametric likelihood ratio inference. *The Annals of Statistics 25*, 1471–1509.

Murphy, S. (1994). Consistency in a proportional hazards model incorporating a random effect. *The Annals of Statistics 22*(2), 712–731.

Murphy, S. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics 23*(1), 182–198.

Murphy, S. (2000). On profile likelihood. *Journal of the American Statistical Association 95*, 449–485.

Nielsen, G., R. Gill, P. Andersen, and T. Sorensen (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics 19*, 25–43.

Peng, Y. and K. Dear (2000). A nonparametric mixture model for cure rate estimation. *Biometrics 56*, 237–243.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling (1994). *Numerical Recipies in Pascal. The Art of Scientific Computing.* New York, NY: Cambridge University Press.

Sy, J. and J. Taylor (2000). Estimation in a Cox proportional hazards cure model. *Biometrics 56*, 227–236.

Tsodikov, A., J. Ibrahim, and A. Yakovlev (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association.*

Tsodikov, A. (2002). Semiparametric models of long- and short-term survival: An application to the analysis of breast cancer survival in Utah by age and stage. *Statistics in Medicine 21*, 895–920.

Tsodikov, A. (2003). Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society, Series B.*

van der Vaart, A. (1998). *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK: Cambridge University Press.

Vaupel, J., K. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography 16*, 439–454.

Wassel, J. and M. Moeschberger (1993). A bivariate survival model with modified gamma frailty for assessing the impact of interventions. *Statistics in Medicine 12*, 241–248.

Lifetime Data Analysis, 10, 103–120, 2004

Two-Sample Statistics for Testing the Equality of Survival Functions Against Improper Semi-parametric Accelerated Failure Time Alternatives: An Application to the Analysis of a Breast Cancer Clinical Trial

PHILIPPE BROËT broet@vjf.inserm.fr
Department of Public Health and National Institute for Health and Medical Research (INSERM U472) Paul Brousse Hospital, Villejuif, France and Institut Curie, 26 rue d'Ulm, Paris, France

ALEXANDER TSODIKOV
Huntsman Cancer Institute 2000 Circle of Hope, Salt Lake City, Utah 84112-5550, USA

YANN DE RYCKE
Institut Curie, 26 rue d'Ulm, Paris, France

THIERRY MOREAU
National Institute for Health and Medical Research (INSERM U472) Paul Brousse Hospital, Villejuif, France

Received October 23, 2002; Revised June 20, 2003; Accepted August 4, 2003

Abstract. This paper presents two-sample statistics suited for testing equality of survival functions against improper semi-parametric accelerated failure time alternatives. These tests are designed for comparing either the short- or the long-term effect of a prognostic factor, or both. These statistics are obtained as partial likelihood score statistics from a time-dependent Cox model. As a consequence, the proposed tests can be very easily implemented using widely available software. A breast cancer clinical trial is presented as an example to demonstrate the utility of the proposed tests.

Keywords: accelerated failure time models, cure rate model, improper model, semi-parametric model

1. Introduction

In recent years, there has been a renewed interest in methods for analyzing survival data with long-term survivors fraction or a 'cure fraction' (for a review, see Maller and Zhou, 1996). Most of these methods attempt to distinguish between the different mechanisms by which a prognostic factor may act on the event's occurrence. Indeed, a prognostic factor may affect either the probability of never experiencing the event of interest (termed 'long-term effect' in the following text) or the time to occurrence of the event (termed 'short-term effect' in the following text), or both.

Testing procedures for these effects were proposed in a recent paper where we assumed proportional hazards for the short-term effect (Broët et al., 2001). In the present paper, we extend this procedure to a non-proportional hazards behavior

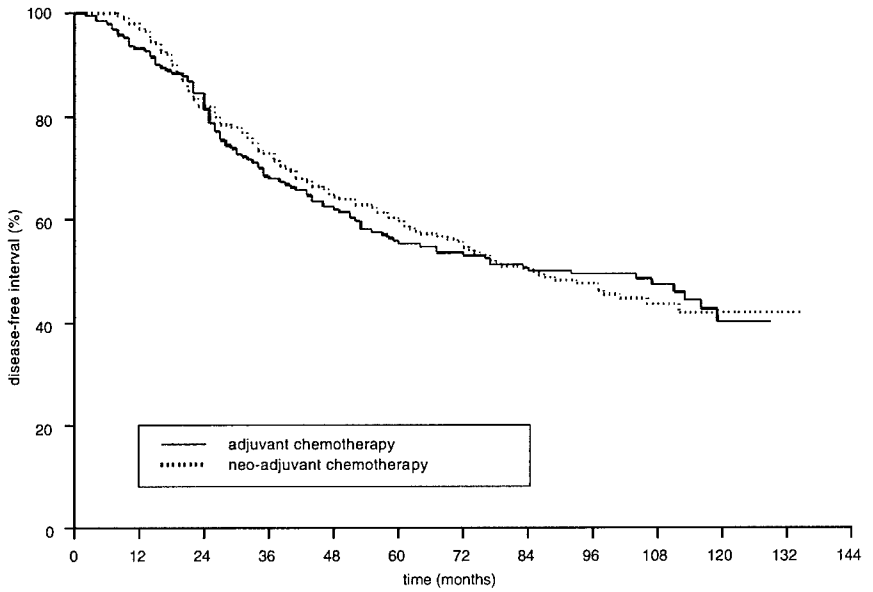


Figure 1. Kaplan–Meier estimates of the recurrence-free interval according to the group of treatment.

of the short-term effect. This work was motivated by the analysis of a breast cancer randomized trial comparing the distribution of the disease-free interval in two treatments groups where Kaplan–Meier curves (Kaplan and Meier, 1958) cross each other during the follow-up (see Figure 1). The two groups are defined according to the different way of administration of the same chemotherapy which is scheduled either to follow (adjuvant) or precede (primary or neo-adjuvant) the local regional treatment. The question addressed in Section 5 is whether primary chemotherapy, which shrinks the tumor before local treatment, modifies the timing of the recurrence as compared to adjuvant chemotherapy, taking a long-term recurrence-free rate into account. This situation was quite puzzling and prompted us to derive test statistics suited for this case. As it will be seen, the proposed tests provide interesting results that would have been overlooked if only results from classical tests were considered.

In the literature, the most common approach for modeling failure-time data with a cure fraction relies on the assumption that the overall distribution of the survival times is a mixture of two components: one corresponding to the subjects who are not susceptible ('cured' subjects) and the other to the subjects who are susceptible of experiencing the event ('uncured' subjects). In this setting, most of the published non-parametric procedures which are convenient for the two-sample comparison do not allow testing for both effects (short-term and long-term) (Gray and Tsiatis, 1989;

Laska and Meisner, 1992; Sposto et al. 1992; Lee, 1995). Others require complex computations which are too heavy for a practical use in routine (Kuk and Chen, 1992; Taylor, 1995).

A different approach which overcomes these drawbacks relies on models that define the cumulative hazard as a bounded increasing positive function in a parametric (Aalen, 1992; Cantor and Shuster, 1992; Yakovlev and Tsodikov, 1996) or semi-parametric way (Tsodikov, 1998; Shen and Sinha, 2002; Ibrahim, 1999, 2001; Broët et al., 2001). Such semi-parametric modeling was used in our previous work to test for no short-term or no long-term effect against improper short-term proportional hazard alternatives (Broët et al., 2001). In this paper, the restrictive proportional hazard short-term effect assumption is relaxed and statistics are proposed for improper short-term accelerated failure time alternatives.

In Section 2, a semi-parametric improper accelerated failure time model is described. In Section 3, the proposed score statistics are derived from a Cox model with a time-dependent covariate. In Section 4, we present the results of simulation experiments. In Section 5, the clinical relevance of these tests is demonstrated by the analysis of a breast cancer clinical trial with long-term follow-up. Section 6 contains a discussion and guidelines for the use of the tests.

2. The Semi-Parametric Improper Accelerated Failure Time Model

Let $i = 0, 1$ denote the two groups to be compared, with n_i subjects in group i ($n = n_0 + n_1$). For each patient j, let the random variables T_j and C_j be the survival and censoring times which are assumed to satisfy the condition of independent censoring (Fleming and Harrington, 1991, pp. 26–27). We denote $X_j = \min(T_j, C_j)$ the observed time of follow-up, $\delta_j = 1_{\{X_j = T_j\}}$ the indicator of death, $Y_j(t) = 1_{\{t \leq X_j\}}$ the indicator of being at risk at time t, and Z_j the indicator variable of group 1. For the subject j, the data consist of X_j, δ_j and Z_j. The hazard function of T_j corresponding to every subject j belonging to group i is denoted by: $\lambda_i(t) = f_i(t)/S_i(t)$, where $f_i(t)$ and $S_i(t)$ are the probability density function and the survival function, respectively. The corresponding cumulative hazard function is denoted by $\Lambda_i(t) = -\log[S_i(t)]$

A semi-parametric improper model is defined by the following general survival function in group i:

$$S_i(t) = \exp\{-\theta e^{\beta_1 i}[1 - A(t, \beta_2 i)]\} \tag{1}$$

where $A(t, \beta_2 i)$ is a function decreasing with time from one to zero, which is similar to a survival function, and where θ is a positive parameter. The function $S_i(t)$ is improper and its limiting value $\exp(-\theta e^{\beta_1 i})$ is called the tail defect and represents the probability of not experiencing the event of interest in group i. The cumulative hazard $\Lambda_i(t) = \theta e^{\beta_1 i}[1 - A(t, \beta_2 i)]$ is less than or equal to $\theta e^{\beta_1 i}$.

The model (1) has two components: the first term containing β_1 which quantifies the long-term effect and the function $A(t, \beta_2 i)$ which expresses the short-term

effect. More precisely, if $\beta_1 = 0$, the two groups have the same cure fraction (no long-term effect) and if $\beta_2 = 0$, the model reduces to a proportional hazard model. In this case, the relative risk is constant over time which implies no short-term effect.

A particular case of (1) was considered in a previous work (Broët et al., 2001), where we assumed a proportional hazard modeling of the short-term effect, so that in the general formulation (1) given here, $A(t, \beta_2 i) = A(t)^{e^{\beta_2 i}}$. Here, we consider another particular case of (1) where a non-proportional hazards model is assumed for the short-term effect by letting: $A(t, \beta_2 i) = \exp\left[-K(t)^{e^{\beta_2 i}}\right]$ where $K(t)$ is a positive function increasing with time from zero to infinity. This is an obvious semi-parametric generalization of the case of two Weibull distributions differing in their shape parameters. The resulting model

$$S_i(t) = \exp\left\{-\theta e^{\beta_1 i}\left[1 - \exp\left(-K(t)^{e^{\beta_2 i}}\right)\right]\right\} \tag{2}$$

has the following property. In case of no long-term effect ($\beta_1 = 0$) and with a short-term effect such as $\beta_2 < 0$, the survival functions $S_0(t)$ and $S_1(t)$ cross before converging to the same long-term survivor fraction.

3. Proposed Test Statistics

In this section, statistics are derived for testing ($\beta_1 = 0$) and/or ($\beta_2 = 0$) in model (2). The derivation is achieved by using a proportional hazards model with a time-dependent covariate which approximates (2) about ($\beta_2 = 0$) and which serves as a basis for computing the desired statistic. They are easily computed as score statistics from the usual partial likelihood. The null hypotheses to be tested are: $H_0: (\beta_1 = \beta_2 = 0)$; $H_{00}: (\beta_2 = 0)$ and $H_{000}: (\beta_1 = 0)$.

3.1. Method for Deriving the Test Statistics

Now, we define the following quantity: $D(t, \beta_2 i) = -\frac{\partial}{\partial t} A(t, \beta_2 i)$ which refers to the density function related to $A(t, \beta_2 i)$. The general model (1) can be written in terms of the hazard functions $\lambda_0(t)$ and $\lambda_1(t)$:

$$\log[\lambda_1(t)/\lambda_0(t)] = \beta_1 + \log[D(t, \beta_2)/D(t, 0)] \tag{3}$$

Expanding $\log\left(D\left(t, \beta_2\right)\right)$ about $\beta_2 = 0$ in (3) gives the following first-order approximation:

$$\log[\lambda_1(t)/\lambda_0(t)] = \beta_1 + \beta_2 w(t) \tag{4}$$

with

$$w(t) = \left[\frac{\partial}{\partial \beta_2} \log D(t, \beta_2)|_{\beta_2 = 0}\right]$$

Under the improper short-term accelerated failure time model (2), $w(t)$ is equal to

$$w(t) = 1 + \log[-\log A(t)] + [\log A(t)]\log[-\log A(t)]$$

where

$$A(t) = \left[1 - \frac{\Lambda_0(t)}{\theta}\right]. \tag{5}$$

In case of improper short-term proportional hazard model, $w(t) = 1 + [\log A(t)]$. Substituting $\Lambda_0(t)$ and θ in (5) by efficient estimators under the null hypothesis to be tested provides estimates $\hat{w}(t)$ of $w(t)$. These estimators are presented for each null hypothesis H_0, H_{00} and H_{000} in the next subsection. Replacing $w(t)$ by $\hat{w}(t)$ in (4), defines a time-dependent proportional hazards model with the internal time-dependent covariate $\hat{w}(t)$ (Kabfleisch and Prentice, 1980). The proposed statistics for testing $(\beta_1 = 0)$ and/or $(\beta_2 = 0)$ can be easily derived as the score statistics from this time-dependent proportional hazards model through the corresponding partial likelihood. It can be easily shown that the resulting score statistics for testing the lack of short-term effect with or without a long-term effect are the same as in model (2), which would not be the case for the likelihood ratio or Wald tests. Moreover, the proposed score statistic for testing for no long-term effect can be easily derived while similar derivation from model (2) would be at least burdensome.

The resulting score statistics depend on the unknown parameters $\Lambda_0(t)$ and θ. Replacing $\Lambda_0(t)$ and θ by efficient estimators and applying the results of Pierce (Pierce, 1982) to our setting as presented in an earlier work (Broët et al., 2001) for improper short-term proportional hazards model allows us to obtain the asymptotic distributions of the proposed statistics.

Score statistics for testing H_0 and H_{00} are asymptotically distributed as chi-squares with two degrees and one degree of freedom, respectively. Concerning H_{000}, it should be noted that the corresponding score statistic depends on an estimate of β_2 as seen in the next section. As the score statistic is derived from model 3 (based on a first-order approximation) which is valid under $\beta_2 = 0$, the score statistic is approximately distributed as a χ^2 with one degree of freedom for small values of β_2. This is not the case for the two other tests that do not depend on β_2 under their corresponding null hypothesis. For the validity of the results it is required that the upper bound of the domain for which the survival distribution of the survival time variable is greater than zero, be less than the upper bound of the censoring distribution. In practice, this condition expresses the fact that the susceptible subjects should experience the event within the maximum length of follow-up. It should be stressed that the distribution of the score statistics for testing H_0 and H_{00} is a chi-square distribution no matter whether the sufficient follow-up condition holds true or not. Indeed, the null hypotheses H_0 and H_{00} do not involve $A(t)$ and are identical under the two models.

3.2. *Score Tests*

3.2.1. *Testing the Lack of Short and Long-Term Effect*

The components of the score vector for testing $H_0 : \beta_1 = \beta_2 = 0$ can be written as follows:

$$\hat{V}_{H_0,1} = \frac{\partial \log L}{\partial \beta_1} = \sum_{j=1}^{n} \delta_j \left\{ Z_j - \frac{\sum_{k=1}^{n} Y_k(t_j) \cdot Z_k}{\sum_{k=1}^{n} Y_k(t_j)} \right\}$$

$$\hat{V}_{H_0,2} = \frac{\partial \log L}{\partial \beta_2} = \sum_{j=1}^{n} \delta_j \hat{w}(t_j) \left\{ Z_j - \frac{\sum_{k=1}^{n} Y_k(t_j) \cdot Z_k}{\sum_{k=1}^{n} Y_k(t_j)} \right\}$$

In $\hat{V}_{H_0,2}$, $\hat{w}(t)$ is computed as indicated in Section 3.1 by using the left-continuous version of the Nelson–Aalen estimator (Nelson, 1972; Aalen, 1978) for $\hat{\Lambda}_0(t)$ and using its value computed at the last observed failure time for $\hat{\theta}$. The corresponding observed information matrix $\hat{I}_{H_0}$ under H_0 is given in the Appendix.

Under H_0, the statistic $S_{H_0} = [\hat{V}_{H_0,1}, \hat{V}_{H_0,2}] \hat{I}_{H_0}^{-1} [\hat{V}_{H_0,1}, \hat{V}_{H_0,2}]'$ is asymptotically distributed as a chi-square with two degrees of freedom.

3.2.2. *Testing the Lack of Short-Term Effect*

The components of the score vector for testing H_{00}: $\beta_2 = 0$, *for any* β_1 can be written as follows:

$$\hat{V}_{H_{00},1} = \frac{\partial \log L}{\partial \beta_1} = 0$$

$$\hat{V}_{H_{00},2} = \frac{\partial \log L}{\partial \beta_2} = \sum_{j=1}^{n} \delta_j \hat{w}(t_j) \left\{ Z_j - \frac{\sum_{k=1}^{n} Y_k(t_j) e^{\hat{\beta}_1 Z_k} Z_k}{\sum_{k=1}^{n} Y_k(t_j) e^{\hat{\beta}_1 Z_k}} \right\}$$

In $\hat{V}_{H_{00},2}$, $\hat{\beta}_1$ is the usual partial likelihood estimator of β_1 under H_{00}; $\hat{w}(t)$ is computed by using the left-continuous version of the Breslow's estimator [Breslow, 1972,1974] for $\hat{\Lambda}_0(t)$ under H_{00} and for $\hat{\theta}$ its value computed at the last observed failure time. The Breslow's estimator for $\Lambda_0(t)$ under H_{00} is given by

$$\sum_{k=1}^{n} \delta_k \left[\sum_{j=1}^{n} Y_j(t_k) \right]^{-1}.$$

The corresponding observed information matrix $\hat{I}_{H_{00}}$ under H_{00} is given in the Appendix.

Under H_{00}, the statistic $S_{H_{00}} = [0, \hat{V}_{H_{00},2}] \hat{I}_{H_{00}}^{-1} [0, \hat{V}_{H_{00},2}]'$ is asymptotically distributed as a χ^2 with one degree of freedom.

3.2.3. *Testing the Lack of Long-Term Effect*

The components of the score vector for testing H_{000}: $\beta_1 = 0$, *for any* β_2 can be written as follows:

$$\hat{V}_{H_{000},1} = \frac{\partial \log L}{\partial \beta_1} = \sum_{j=1}^{n} \delta_j \left\{ Z_j - \frac{\sum_{k=1}^{n} Y_k(t_j) e^{\hat{\beta}_2 \hat{w}(t_j) Z_k} Z_k}{\sum_{k=1}^{n} Y_k(t_j) e^{\hat{\beta}_2 \hat{w}(t_j) Z_k}} \right\}$$

$$\hat{V}_{H_{000},2} = \frac{\partial \log L}{\partial \beta_2} = 0$$

where $\hat{V}_{H_{000},1}$ is obtained by using $\hat{w}(t)$ as given in Section 3.2.

A derivation of $\hat{w}(t)$ and $\hat{\beta}_2$ could be achieved through an iterative procedure. For computational simplicity, the first-step estimators are used in the proposed score statistic. The procedure is as follows. At first step, $\hat{w}(t)$ is taken under H_0 where the cumulative baseline hazard is replaced by the Nelson–Aalen estimator and $\hat{\beta}_2$ is taken as the partial likelihood estimator obtained in the corresponding time-dependent model. At second step, $\hat{\beta}_2$ thus obtained, is used to update $\hat{w}(t)$ usingthe left-continuous version of a Breslow's type estimator given by

$$\sum_{k=1}^{n} \delta_k \left[\sum_{j=1}^{n} Y_j(t_k) e^{\left[Z_j \hat{\beta}_2 \hat{w}(t_k) \right]} \right]^{-1}$$

for $\hat{\Lambda}_0(t)$. As mentioned above, the proposed statistic is computed by using the first-step estimator $\hat{\beta}_2$ together with $\hat{w}(t)$. This implies that $\hat{V}_{H_{000},2}$ does not vanish but is taken to be null in the computation. The corresponding observed information matrix $\hat{I}_{H_{000}}$ is given in the Appendix.

For small values of β_2, under H_{000}, the statistic $S_{H_{000}} = \left[\hat{V}_{H_{000},1}, 0 \right] \hat{I}^{-1}_{H_{000}} \left[\hat{V}_{H_{000},1}, 0 \right]'$ is approximately distributed as a χ^2 with one degree of freedom.

4. Simulation Study

4.1. Method

A simulation study was performed to investigate the power properties of the proposed tests in comparison with classical tests such as the Logrank test (LR) (Peto and Peto, 1972) and the Peto–Prentice–Wilcoxon test (PPW) (Kabfleisch and Prentice, 1980). The proposed tests of H_0, H_{00} and H_{000} are denoted SLT, ST and LT, respectively. We also consider the test for no short- and no long-term effect (SLT-PH) designed for improper short-term proportional hazard alternatives [Broët et al., 2001]. In addition, the product-limit test (PL) which is a non-parametric test of no difference in the cure fraction (Sposto, Sather and Baker, 1992) is also considered.

Data were generated to mimic a simple randomized clinical trial with two different models: (A) improper short-term accelerated failure time model, (B) improper short-term proportional hazard model. Survival times were generated according to model (1) with $A(t, \beta_2 i) = \exp\left(-te^{\beta_2 i}\right)$ and $A(t, \beta_2 i) = \exp\left(-t^{e^{\beta_2 i}}\right)$ for the proportional

hazards and accelerated failure time model, respectively. Censoring times were independently generated from a uniform distribution over $[0, u]$. It is worth noting that in the uniform censoring case, a susceptible subject may not experience the event of interest within the follow-up time u. For each set of parameter values u can be easily computed so as to ensure a given percentage of censoring. The percentage of censoring refers only to the percentage of censored observations without the cure fraction $\exp(-\theta)$. The number of subjects per group was chosen to be 100. The following configurations were considered: $\exp(-\theta) = 0.3, 0.5, 0.7$; 0%, 20% and 40% censoring; $e^{\beta_1} = 2/3, 1, 3/2$ and $e^{\beta_2} = 0.5, 1, 2$. For 20% censoring as specified above, the actual rate of censoring was 44%, 60% and 76% for each plateau value. For 40% censoring, the actual rate of censoring was 58%, 70% and 82% for each plateau value. For each configuration, 1,000 replications were performed and the levels and powers of all tests were estimated at the nominal level of 0.05.

4.2. Results

Tables 1(a–c) display the results for model (A) whereas Tables 2(a–c) display those for model (B).

Table 1(a) shows the results obtained in the uncensored case. Except for the LT test, the estimated level of each test under its proper null hypothesis is within the binomial range [0.036; 0.064]. In the presence of a short-term effect, the observed levels for the LT test increase up to 10%. The test of no short-term and long-term effect (SLT) shows a strongly increased power relative to LR, PPW and SLT-PH in the presence of a short-term effect. The power gains are striking for no, or small differences in the long-term effects. As compared to LR, the SLT test is in some cases 10 times more powerful. However, it is well known that the LR test is not suited for such situations where survival curves cross. In case of no difference in short-term effects, the power of this latter test is slightly decreased relative to that of the LR. In any case it is less than 12% lower than that of the LR test. Power values of the ST test are very close to the SLT. Regarding the long-term effect, the PL test is more powerful than LT and LR. Power of these two latter tests is quite close.

Table 1(b) shows the results obtained with a 20% censoring rate. The observed levels of the SLT and ST tests do not exceed the binomial bounds. This is not the case for LT and PL where the observed level is increased up to 9% in case a short-term effect exists. Concerning the power, it appears that the trends observed in the uncensored case remain almost unchanged. Power gains for the ST and SLT relative to LR are lower than in the uncensored case, but still remain impressive as compared to LR. For the LT, the magnitude of the power values is lower than in the uncensored case and is always under those of the PL test.

The results obtained at a 40% censoring rate are shown in table 1(c). Regarding the SLT and ST tests, empirical significance levels appear to be close to the nominal level, and power gains are less pronounced than at lower censoring rates. However, with a short-term effect, the magnitude of the power gain remains high. Concerning the LT test, the observed levels are appreciably higher than the nominal level in the

Table 1. Simulation results for the improper short-term accelerated failure time model with (a) no censoring, (b) 20% censoring and (c) 40% censoring.

$e^{-\theta}$	e^{β_1}	e^{β_2}	LR	PPW	SLT-PH	SLT	ST	LT	PL
(a) No censoring									
		0.5	0.482	0.219	0.362	0.993	0.995	0.478	0.598
	0.67	1.0	0.648	0.613	0.532	0.530	0.062	0.649	0.593
		2.0	0.757	0.873	0.657	0.999	0.992	0.787	0.575
		0.5	0.083	0.268	0.088	0.992	0.999	0.089	0.052
30%	1.00	1.0	0.043	0.046	0.048	0.048	0.056	0.040	0.046
		2.0	0.096	0.284	0.097	0.993	0.998	0.092	0.057
		0.5	0.901	0.988	0.857	1.000	1.000	0.858	0.626
	1.50	1.0	0.710	0.650	0.615	0.601	0.041	0.711	0.626
		2.0	0.411	0.099	0.407	1.000	1.000	0.557	0.657
		0.5	0.409	0.338	0.347	0.960	0.950	0.359	0.476
	0.67	1.0	0.466	0.445	0.379	0.366	0.056	0.470	0.472
		2.0	0.499	0.596	0.376	0.973	0.949	0.625	0.436
		0.5	0.052	0.070	0.050	0.957	0.978	0.073	0.049
50%	1.00	1.0	0.036	0.039	0.034	0.051	0.054	0.034	0.044
		2.0	0.049	0.088	0.054	0.950	0.972	0.094	0.048
		0.5	0.661	0.791	0.562	0.998	0.989	0.730	0.538
	1.50	1.0	0.575	0.563	0.471	0.476	0.063	0.579	0.569
		2.0	0.474	0.268	0.367	0.994	0.994	0.452	0.586
		0.5	0.295	0.265	0.230	0.754	0.751	0.240	0.309
	0.67	1.0	0.314	0.308	0.231	0.241	0.057	0.321	0.317
		2.0	0.320	0.340	0.230	0.771	0.721	0.468	0.302
		0.5	0.055	0.056	0.044	0.719	0.831	0.093	0.048
70%	1.00	1.0	0.047	0.047	0.040	0.047	0.048	0.047	0.046
		2.0	0.049	0.053	0.048	0.748	0.839	0.081	0.051
		0.5	0.440	0.479	0.348	0.925	0.884	0.571	0.401
	1.50	1.0	0.405	0.405	0.307	0.295	0.039	0.405	0.391
		2.0	0.398	0.335	0.299	0.913	0.906	0.332	0.435
(b) 20% censoring									
		0.5	0.518	0.234	0.683	0.980	0.984	0.530	0.615
	0.67	1.0	0.570	0.548	0.464	0.467	0.057	0.574	0.460
		2.0	0.722	0.869	0.663	0.997	0.990	0.737	0.499
		0.5	0.079	0.322	0.683	0.972	0.979	0.084	0.106
30%	1.00	1.0	0.053	0.051	0.063	0.055	0.057	0.054	0.058
		2.0	0.102	0.297	0.371	0.981	0.991	0.075	0.053
		0.5	0.908	0.988	0.986	1.000	0.986	0.845	0.204
	1.50	1.0	0.569	0.523	0.479	0.422	0.051	0.563	0.348
		2.0	0.184	0.039	0.730	0.957	0.991	0.451	0.492
		0.5	0.438	0.305	0.434	0.905	0.902	0.377	0.475
	0.67	1.0	0.424	0.416	0.348	0.342	0.049	0.432	0.429
		2.0	0.500	0.574	0.393	0.949	0.904	0.618	0.407
		0.5	0.049	0.095	0.415	0.850	0.927	0.082	0.087

Table 1. Continued.

$e^{-\theta}$	e^{β_1}	e^{β_2}	LR	PPW	SLT-PH	SLT	ST	LT	PL
50%	1.00	1.0	0.048	0.049	0.063	0.056	0.063	0.049	0.049
		2.0	0.060	0.086	0.142	0.860	0.933	0.079	0.061
		0.5	0.680	0.820	0.863	0.973	0.908	0.703	0.162
	1.50	1.0	0.405	0.382	0.315	0.305	0.059	0.400	0.298
		2.0	0.298	0.128	0.577	0.890	0.948	0.398	0.475
		0.5	0.326	0.293	0.261	0.693	0.662	0.269	0.311
	0.67	1.0	0.275	0.274	0.217	0.202	0.047	0.287	0.279
		2.0	0.287	0.318	0.211	0.716	0.661	0.421	0.275
		0.5	0.045	0.038	0.192	0.544	0.674	0.055	0.071
70%	1.00	1.0	0.053	0.057	0.055	0.057	0.052	0.057	0.057
		2.0	0.044	0.055	0.096	0.560	0.658	0.071	0.046
		0.5	0.342	0.414	0.560	0.767	0.649	0.436	0.138
	1.50	1.0	0.274	0.279	0.214	0.213	0.050	0.286	0.228
		2.0	0.244	0.183	0.375	0.632	0.734	0.276	0.342
(c) 40% censoring									
		0.5	0.198	0.084	0.707	0.892	0.954	0.270	0.474
	0.67	1.0	0.474	0.457	0.393	0.379	0.047	0.475	0.318
		2.0	0.704	0.869	0.868	0.987	0.940	0.610	0.158
		0.5	0.303	0.535	0.790	0.948	0.950	0.283	0.056
30%	1.00	1.0	0.050	0.054	0.061	0.046	0.050	0.054	0.050
		2.0	0.197	0.447	0.727	0.948	0.966	0.110	0.069
		0.5	0.976	0.992	0.991	1.000	0.911	0.947	0.407
	1.50	1.0	0.500	0.473	0.394	0.379	0.055	0.485	0.295
		2.0	0.052	0.097	0.719	0.911	0.984	0.185	0.322
		0.5	0.243	0.160	0.557	0.747	0.839	0.249	0.434
	0.67	1.0	0.327	0.316	0.239	0.236	0.042	0.328	0.246
		2.0	0.425	0.553	0.555	0.880	0.816	0.466	0.173
		0.5	0.093	0.150	0.514	0.772	0.825	0.132	0.044
50%	1.00	1.0	0.054	0.059	0.052	0.047	0.050	0.061	0.053
		2.0	0.083	0.144	0.475	0.770	0.843	0.083	0.081
		0.5	0.807	0.890	0.923	0.973	0.749	0.817	0.277
	1.50	1.0	0.341	0.328	0.259	0.257	0.056	0.336	0.230
		2.0	0.091	0.057	0.552	0.734	0.896	0.207	0.354
		0.5	0.229	0.187	0.378	0.503	0.582	0.229	0.307
	0.67	1.0	0.213	0.217	0.161	0.172	0.054	0.224	0.189
		2.0	0.213	0.243	0.230	0.538	0.473	0.282	0.130
		0.5	0.046	0.056	0.287	0.443	0.574	0.072	0.054
70%	1.00	1.0	0.040	0.040	0.047	0.037	0.044	0.043	0.044
		2.0	0.050	0.058	0.231	0.468	0.596	0.069	0.062
		0.5	0.504	0.557	0.643	0.773	0.468	0.563	0.191
	1.50	1.0	0.229	0.228	0.150	0.157	0.045	0.236	0.172
		2.0	0.120	0.085	0.382	0.472	0.661	0.193	0.293

Configurations presented below correspond to different short-term effect (e^{β_2}), long-term effect (e^{β_1}) and plateau values ($e^{-\theta}$). The total number of subjects is 200.

Table 2. Simulation results for the improper short-term proportional hazard model with (a) no censoring, (b) 20% censoring and (c) 40% censoring.

$e^{-\theta}$	e^{β_1}	e^{β_2}	LR	PPW	SLT-PH	SLT	ST	LT	PL
(a) No censoring									
		0.5	0.887	0.963	0.989	0.823	0.888	0.882	0.601
	0.67	1.0	0.609	0.591	0.517	0.519	0.048	0.611	0.579
		2.0	0.303	0.111	0.880	0.220	0.895	0.311	0.568
		0.5	0.207	0.450	0.870	0.186	0.894	0.196	0.046
30%	1.00	1.0	0.050	0.048	0.046	0.050	0.041	0.051	0.061
		2.0	0.209	0.487	0.896	0.193	0.916	0.186	0.039
		0.5	0.208	0.057	0.889	0.150	0.909	0.225	0.600
	1.50	1.0	0.711	0.674	0.600	0.592	0.038	0.692	0.609
		2.0	0.972	0.994	0.997	0.956	0.873	0.963	0.646
		0.5	0.624	0.758	0.922	0.503	0.812	0.624	0.463
	0.67	1.0	0.480	0.475	0.393	0.381	0.047	0.480	0.487
		2.0	0.310	0.188	0.800	0.221	0.810	0.314	0.444
		0.5	0.084	0.192	0.783	0.062	0.854	0.084	0.046
50%	1.00	1.0	0.048	0.048	0.043	0.056	0.052	0.048	0.051
		2.0	0.087	0.188	0.784	0.065	0.853	0.085	0.051
		0.5	0.310	0.146	0.878	0.224	0.882	0.314	0.573
	1.50	1.0	0.579	0.574	0.490	0.488	0.061	0.581	0.569
		2.0	0.847	0.935	0.978	0.739	0.866	0.840	0.567
		0.5	0.353	0.418	0.686	0.253	0.606	0.354	0.293
	0.67	1.0	0.322	0.327	0.252	0.239	0.047	0.328	0.317
		2.0	0.257	0.200	0.570	0.176	0.543	0.255	0.311
		0.5	0.053	0.073	0.540	0.043	0.660	0.053	0.054
70%	1.00	1.0	0.047	0.047	0.047	0.051	0.042	0.048	0.053
		2.0	0.066	0.086	0.562	0.047	0.681	0.067	0.059
		0.5	0.291	0.206	0.763	0.216	0.713	0.292	0.402
	1.50	1.0	0.405	0.394	0.300	0.305	0.044	0.409	0.398
		2.0	0.505	0.602	0.841	0.381	0.739	0.507	0.387
(b) 20% censoring									
		0.5	0.928	0.971	0.980	0.909	0.686	0.920	0.496
	0.67	1.0	0.559	0.534	0.456	0.447	0.049	0.555	0.476
		2.0	0.146	0.055	0.677	0.153	0.720	0.161	0.463
		0.5	0.404	0.561	0.623	0.476	0.462	0.358	0.069
30%	1.00	1.0	0.042	0.038	0.044	0.040	0.053	0.035	0.046
		2.0	0.438	0.584	0.636	0.482	0.461	0.391	0.058
		0.5	0.039	0.066	0.207	0.084	0.286	0.041	0.106
	1.50	1.0	0.563	0.531	0.466	0.429	0.049	0.546	0.314
		2.0	0.989	0.994	0.987	0.986	0.238	0.981	0.590
		0.5	0.699	0.787	0.891	0.615	0.677	0.706	0.407
	0.67	1.0	0.421	0.403	0.319	0.315	0.050	0.423	0.398
		2.0	0.207	0.130	0.652	0.165	0.665	0.207	0.416
		0.5	0.236	0.316	0.458	0.283	0.420	0.232	0.055

Table 2. Continued.

$e^{-\theta}$	e^{β_1}	e^{β_2}	LR	PPW	SLT-PH	SLT	ST	LT	PL
50%	1.00	1.0	0.049	0.050	0.050	0.045	0.050	0.047	0.052
		2.0	0.205	0.288	0.467	0.249	0.447	0.205	0.057
		0.5	0.058	0.057	0.188	0.079	0.235	0.066	0.125
	1.50	1.0	0.429	0.416	0.343	0.329	0.051	0.431	0.297
		2.0	0.932	0.946	0.916	0.909	0.234	0.918	0.557
		0.5	0.405	0.460	0.654	0.317	0.529	0.421	0.286
	0.67	1.0	0.296	0.290	0.225	0.226	0.043	0.304	0.293
		2.0	0.213	0.179	0.470	0.170	0.458	0.207	0.296
		0.5	0.106	0.133	0.289	0.138	0.326	0.115	0.050
70%	1.00	1.0	0.052	0.052	0.050	0.045	0.044	0.056	0.054
		2.0	0.111	0.138	0.299	0.146	0.321	0.126	0.067
		0.5	0.054	0.048	0.152	0.086	0.183	0.059	0.095
	1.50	1.0	0.289	0.286	0.213	0.217	0.042	0.296	0.229
		2.0	0.738	0.768	0.698	0.673	0.185	0.740	0.430
(c) 40% censoring									
		0.5	0.959	0.971	0.946	0.943	0.221	0.948	0.587
	0.67	1.0	0.448	0.421	0.345	0.335	0.036	0.456	0.292
		2.0	0.052	0.049	0.169	0.079	0.222	0.051	0.079
		0.5	0.573	0.624	0.527	0.507	0.143	0.538	0.209
30%	1.00	1.0	0.049	0.050	0.060	0.048	0.051	0.055	0.053
		2.0	0.559	0.618	0.523	0.525	0.141	0.527	0.184
		0.5	0.083	0.085	0.100	0.087	0.094	0.070	0.043
	1.50	1.0	0.447	0.427	0.359	0.338	0.050	0.440	0.244
		2.0	0.990	0.988	0.976	0.972	0.086	0.976	0.700
		0.5	0.819	0.847	0.786	0.763	0.203	0.815	0.463
	0.67	1.0	0.326	0.328	0.257	0.246	0.053	0.332	0.247
		2.0	0.077	0.063	0.177	0.073	0.219	0.083	0.125
		0.5	0.364	0.408	0.345	0.342	0.142	0.351	0.143
50%	1.00	1.0	0.047	0.039	0.039	0.034	0.044	0.050	0.036
		2.0	0.373	0.410	0.368	0.342	0.143	0.358	0.147
		0.5	0.057	0.064	0.097	0.067	0.108	0.052	0.051
	1.50	1.0	0.330	0.319	0.269	0.254	0.060	0.329	0.231
		2.0	0.919	0.928	0.883	0.873	0.097	0.911	0.632
		0.5	0.547	0.572	0.505	0.469	0.172	0.555	0.333
	0.67	1.0	0.213	0.212	0.171	0.169	0.042	0.225	0.201
		2.0	0.058	0.049	0.125	0.076	0.184	0.063	0.122
		0.5	0.178	0.208	0.219	0.193	0.146	0.193	0.114
70%	1.00	1.0	0.062	0.058	0.042	0.051	0.044	0.066	0.055
		2.0	0.225	0.239	0.208	0.185	0.127	0.231	0.111
		0.5	0.064	0.063	0.071	0.064	0.074	0.064	0.059
	1.50	1.0	0.206	0.211	0.143	0.164	0.056	0.206	0.158
		2.0	0.669	0.679	0.591	0.596	0.079	0.666	0.459

Configurations presented below correspond to different short-term effect (e^{β_2}), long-term effect (e^{β_1}) and plateau values ($e^{-\theta}$). The total number of subjects is 200.

presence of a short-term effect. For the PL test, the observed type I error rate is slightly increased up to 8% in case of an existing non-null short-term effect and power gains are less than those observed at a lower censoring rates.

Tables 2(a–c), show the results for model (B). Estimated type I error was very close to the nominal significance level of 0.05 for the SLT and the ST test in every configuration. This is not the case for the LT test which always yields higher observed levels than the nominal level with values markedly increased in some situations. To a lesser extent, a similar trend is observed for the PL test for which observed level is increased up to 18% in case of an existing non-null short-term effect and a high censoring rate. It is worth noting that in this short-term proportional hazards situation, the loss of power of the SLT test remains small as compared to the LR.

Concerning the type I error of the proposed tests, it should be stressed that the null hypotheses H_0 and H_{00} involve neither $\hat{0}$ nor $\hat{w}(t)$. As a result, the SLT and ST tests maintain a correct type I error which is not the case for the LT test under the corresponding null hypothesis when β_2 is not null and the model is not the correct one. In the case of no short-term effect where the estimated level is close to the nominal one it appears that the power of the LT test is not dramatically decreasing as compared to the other tests even if in this case the uniform censoring is likely to hinder the long-term effect.

We performed additional simulations with small sample size (not shown here) and, as expected, it leads to a decrease in power which is more pronounced with a high censoring rate.

5. Application

In this section, we consider a clinical trial on breast cancer disease.

5.1. Primary Chemotherapy Trial

The aim of the present analysis was to investigate short-term and long-term effects of primary chemotherapy on disease recurrence by the proposed tests in a mature trial with more than ten years of follow- up. The so-called 'S6-trial' (Scholl et al., 1994) was conducted to assess whether primary chemotherapy improved survival, as compared to the same chemotherapy scheduled to follow the local regional treatment (adjuvant chemotherapy). Premenopausal breast cancer patients were included between October 1986 and June 1990, and randomized to receive either primary or adjuvant chemotherapy. The criteria for inclusion were as follows: non-metastatic operable breast tumors, largest tumor diameter between 3 and 7 cm, axillary lymph nodes not involved clinically, or involved but not adherent, no prior cancer, no serious concomitant illness. Bilateral, inflammatory or locally advanced breast cancers were not eligible. Two hundred breast cancer patients received primary chemotherapy and 190 adjuvant chemotherapy. Chemotherapy was started either

after completion of the initial assessment (primary) or within 2 weeks of ending the local regional therapy (adjuvant). It consisted of four monthly cycles of intravenous cyclophosphamide, doxorubicin and 5-fluorouracil. Following random assignment to primary or adjuvant chemotherapy, patients were reviewed every 3 months for a year, then every 6 months during the first 5 years following the treatment and at least annually thereafter.

In what follows, we focus on the recurrence-free interval and not on the overall survival which was considered in a previous paper (Scholl et al., 1994). The recurrence-free interval is defined as the time from randomization until progression on the first observation of tumor recurrence (local, regional, distant).

5.2. Results

The median follow-up was 105 months. The 5-year recurrence-free interval rates were 60% [53–67] for patients treated with primary chemotherapy and 55% [48–63] for those treated with adjuvant chemotherapy. The 10-year survival rates were 40% [32–51] for patients treated with primary chemotherapy and 42% [35–50] for those treated with adjuvant chemotherapy. At the end of follow-up and for the 390 patients under study, 208 patients experienced a recurrence of the disease.

Figure 1 displays the Kaplan–Meier estimates of the recurrence-free interval by the treatment group. It shows a plateau value (i.e. long-term fraction) in the survival curves after 10 years, which is not surprising since most of the local and distant recurrences occur in the first decade (Bland and Copeland, 1998). Thus, an improper model appears well suited for these data.

Figure 2 displays the estimated survival function $A_i(t)$. The empirical estimate for $A_i(t) = [1 - (\Lambda_i(t)/\theta_i)]$ in each group is obtained by replacing $\Lambda_i(t)$ and θ_i by the Nelson–Aalen estimator and its value at the last observed failure time, respectively. This plot provides an informal assessment of the proportional hazards hypothesis for the short-term effect. It appears that the two survival functions cross, clearly indicating a non-proportional short-term effect.

In what follows, we present the results of the proposed statistics together with those obtained with the classical logrank statistic and the Peto–Prentice–Wilcoxon. We also provide the results of the SLT-PH test.

When testing for differences in recurrence-free interval, the logrank test ($\chi_1^2 < 0.0001, P = 0.99$) the PPW ($\chi_1^2 = 0.10, P = 0.78$) and the SLT-PH tests ($\chi_2^2 = 0.59, P = 0.75$) are not significant. When testing for an overall effect with an accelerated failure time short-term effect, the SLT test is close to the significance ($\chi_2^2 = 4.14, P = 0.13$). When testing for a short-term effect, the ST test is significant ($\chi_1^2 = 4.13, P = 0.04$). No short-term effect was detected when using the LT test ($\chi_1^2 < 0.01, P = 0.94$) or the plateau test ($PL : \chi_1^2 = 0.08, P = 0.77$). These latter results agree with figure 2 which indicates a non-constant short-term effect. From these results, the disease's recurrences have been significantly delayed by primary chemotherapy but without a benefit on long-term recurrence rate as compared to classical adjuvant chemotherapy.

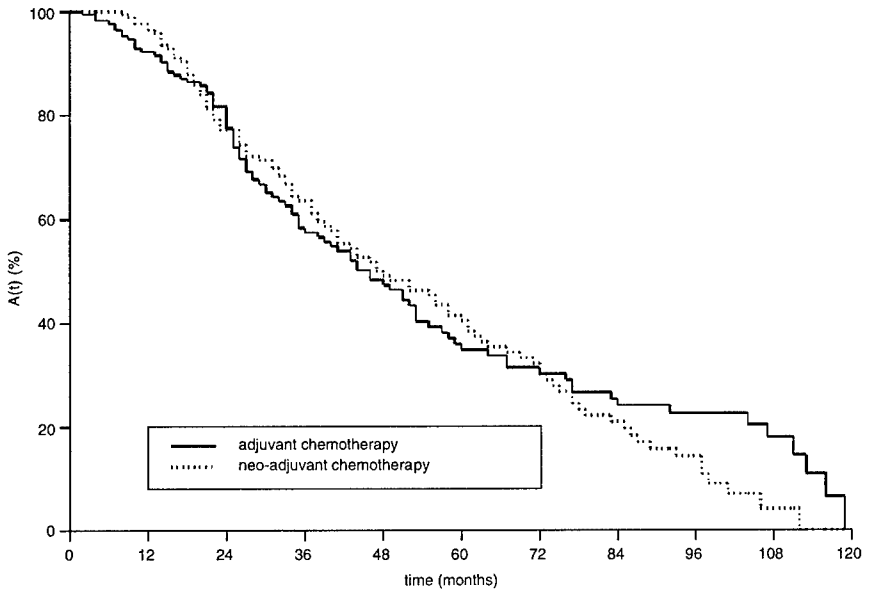


Figure 2. Estimated survival function $A(t)$ according to the group of treatment.

6. Discussion

Survival data with long-term survivors requires extension of existing test statistics for analyzing short and long-term effects of a prognostic factor. In this paper, we propose new score tests well suited for different types of departures from equality of survival distributions with long-term survivors. These tests are related to improper short-term accelerated failure time alternatives and are obtained as score statistics from a time-dependent Cox model.

An interesting feature of these tests is that they are simple to use since they can be very easily obtained from standard Cox model procedures implemented in most statistical software packages. The test of no long-term effect should be particularized since its limiting distribution is obtained in the presence of a negligible short-term effect. This drawback would not exist if a test was derived from the original model, but this does not seem to be computationally realistic in usual practice. It must be kept in mind that using this test also requires that the maximum value of the cumulative hazard be estimated consistently. The theoretical condition underlying this assumption is that sufficient number of patients should be followed up to a time after which the risk of developing the event of interest is negligible. Such drawbacks do not concern the two other tests.

Simulation results show that SLT and ST tests maintain a correct type I error in case of high censoring rate and a misspecified model. In contrast, the proposed LT test is very sensitive to model misspecification and high censoring rates in the presence of a short-term effect. Regarding the power, the SLT and ST show interesting power performances for assessing a short-term effect with or without a long-term effect as compared to classical tests such as the logrank test or the Peto–Prentice–Wilcoxon test. Power gains decrease with censoring which could be explained by the fact that the cumulative hazard is not consistently estimated under the alternative hypothesis. Indeed, the presence of uniform censoring yields to a violation of the sufficient follow-up condition even if the model is correctly specified under the alternative hypothesis.

In practice, SLT and ST tests could be recommended for routine use when a non-constant short-term effect is expected. This could be the case when comparing treatments that modify the speed of progression of the disease in a population where a long-term survivor fraction is commonly encountered. As seen from the simulation results, it seems that with moderate censoring the product-limit test is a more reasonable alternative to the long-term effect tests when a short-term effect is expected.

The proposed score tests are particularly well suited to the study presented in this article since a large proportion of the patients will never recur from the disease and a long-term follow-up is provided. According to our analysis, the recurrence of the disease appeared to have been significantly delayed by primary chemotherapy but without a benefit on long-term recurrence rate as compared to the post-operative chemotherapy. It is tempting to speculate that early and effective targeting of active micrometastasic disease may have delayed the occurrence of disease recurrence. Based on these results, we should emphasize that using the proposed score tests provide some interesting findings for primary chemotherapy that would have been overlooked by only considering results from the classical logrank test. Moreover, we are able to attribute this difference to the short-term effect. Finally, our approach offers a new insight on the different aspects of treatment effects and may be recommended for widespread use in long-term survival studies.

In addition, it should be noted that these tests can obviously be extended for taking into account other factors by using a stratified time-dependent Cox model. Further works are ongoing to extend this family for taking into account other complex time-varying short-term patterns.

In conclusion, the proposed tests, which are very simple to use, seem particularly appealing when testing time-related effects of new markers or therapies in censored data with long-term survivors.

Acknowledgments

The authors would like to acknowledge Prof. A. Yakovlev (University of Utah) and Dr. B. Asselain (Institut Curie) for stimulating discussions during the evolution of

this article. We thank all the members of the Breast Cancer Group of the Curie Institute for their cooperation. We are indebted to Mrs. C. Gautier for her kind technical assistance in carrying out the data management of the S6-trial. This work was supported in part by the 'Société Française du Cancer', NIH/NCI U01 CA 97414 and DAMD 17-03-1-0034.

Appendix: Partial Second Derivatives Based Upon the Working Model

For the following derivation it is convenient to introduce the notations :

$$S^{(0)}[\beta_1,\beta_2,\hat{w}(t),t]=\frac{1}{n}\sum_{k=1}^{n}Y_k(t)e^{\beta_1 Z_k}\exp\{Z_k\beta_2[\hat{w}(t)]\}$$

$$S^{(1)}[\beta_1,\beta_2,\hat{w}(t),t]=\frac{1}{n}\sum_{k=1}^{n}Z_kY_k(t)e^{\beta_1 Z_k}\exp\{Z_k\beta_2[\hat{w}(t)]\}$$

$$S^{(2)}[\beta_1,\beta_2,\hat{w}(t),t]=\frac{1}{n}\sum_{k=1}^{n}Z_k^2Y_k(t)e^{\beta_1 Z_i}\exp\{Z_k\beta_2[\hat{w}(t)]\}$$

It follows that the partial second derivatives are as follows :

$$\frac{\partial^2\log L}{\partial\beta_1\partial\beta_1}=\sum_{j=1}^{n}\delta_j\left\{\left\{\frac{S^{(1)}[\beta_1,\beta_2,\hat{w}(t_j),t_j]}{S^{(0)}[\beta_1,\beta_2,\hat{w}(t),t_j]}\right\}^2-\frac{S^{(2)}[\beta_1,\beta_2,\hat{w}(t_j),t_j]}{S^{(0)}[\beta_1,\beta_2,\hat{w}(t_j),t_j]}\right\}$$

$$\frac{\partial^2\log L}{\partial\beta_2\partial\beta_2}=\sum_{j=1}^{n}\delta_j[\hat{w}(t_j)]^2\left\{\left\{\frac{S^{(1)}[\beta_1,\beta_2,\hat{w}(t_j),t_j]}{S^{(0)}[\beta_1,\beta_2,\hat{w}(t),t_j]}\right\}^2-\frac{S^{(2)}[\beta_1,\beta_2,\hat{w}(t_j),t_j]}{S^{(0)}[\beta_1,\beta_2,\hat{w}(t_j),t_j]}\right\}$$

$$\frac{\partial^2\log L}{\partial\beta_2\partial\beta_1}=\sum_{j=1}^{n}\delta_j[\hat{w}(t_j)]\left\{\left\{\frac{S^{(1)}[\beta_1,\beta_2,\hat{w}(t_j),t_j]}{S^{(0)}[\beta_1,\beta_2,\hat{w}(t),t_j]}\right\}^2-\frac{S^{(2)}[\beta_1,\beta_2,\hat{w}(t_j),t_j]}{S^{(0)}[\beta_1,\beta_2,\hat{w}(t_j),t_j]}\right\}$$

The elements of the information matrix are computed under the null hypothesis H_0 by using $\hat{w}(t)$ as given in Section 3.1 and replacing $S^{(l)}(\beta_1,\beta_2,\hat{w}(t),t)$ by $S^{(l)}(0,0,\hat{w}(t),t)$.

Under the null hypothesis H_{00}, the corresponding elements are computed by using $\hat{w}(t)$ as given in Section 3.1 and replacing $S^{(l)}(\beta_1,\beta_2,\hat{w}(t),t)$ by $S^{(l)}(\hat{\beta}_1,0,\hat{w}(t),t)$.

Finally, under the null hypothesis H_{000}, the corresponding elements are computed by using $\hat{w}(t)$ as given in Section 3.1 and replacing $S^{(l)}(\beta_1,\beta_2,\hat{w}(t),t)$ by $S^{(l)}(0,\hat{\beta}_2,\hat{w}(t),t)$.

References

O. O. Aalen, "Nonparametric inference for a family of counting processes," *Ann. Stat.* vol. 6 pp. 701–726, 1978.

O. O. Aalen, "Modelling heterogeneity in survival analysis by the compound Poisson distribution," *Ann. Appl. Prob.* vol. 2 pp. 951–972, 1992.

K. I. Bland and E. M. Copeland, *The Breast: Comprehensive Management of Benign and Malignant Diseases*. 2nd edn, Saunders: Philadelphia, 1998.

N. E. Breslow, "Contribution to the discussion on the paper by D.R. Cox' Regression and life tables," *J. Roy. Statist. Soc., B* vol. 34 pp. 216–217, 1972.

N. E. Breslow, "Covariance analysis of censored survival datas," *Biometrics* vol. 30 pp. 89–99, 1974.

P. Broët, Y. De Rycke, P. Tubert-Bitter, J. Lellouch, B. Asselain, and T. Moreau, "A semi-parametric approach for the two-sample comparison of survival times with long-term survivors," *Biometrics* vol. 57 pp. 844–852, 2001.

A. B. Cantor and J. J. Shuster, "Parametric versus non-parametric methods for estimating cure rates based on censored survival data," *Stat. Med.* vol. 11 pp. 931–937, 1992.

T. R. Fleming and D. P. Harrington, *Counting Processes and Survival Analysis*, John Wiley, New York, 1991.

R. J. Gray and A. A. Tsiatis, "A linear rank test for use when the main interest is in differences in cure rates," *Biometrics* vol. 45 pp. 899–904, 1989.

J. G. Ibrahim M. H. Chen, and D. Sinha, "A new Bayesian model for survival data with a surviving fraction," *J. Amer. Statist. Assoc.* vol. 94 pp. 909–919, 1999.

J. G. Ibrahim M. H. Chen, and D. Sinha, "Bayesian semiparametric models for survival data with a cure fraction," *Biometrics* vol. 57 pp. 383–388, 2001.

J. D. Kabfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley: New York, 1980.

E. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Am. Stat. Assoc.* vol. 53 pp. 457–581, 1958.

A. Y. C. Kuk and C.-W. Chen, "A mixture model combining logistic regression with proportional hazards regression," *Biometrika* vol. 79 pp. 531–541, 1992.

E. M. Laska and M. J. Meisner, "Nonparametric estimator and testing in a cure models," *Biometrics* vol. 48 pp. 1223–1234, 1992.

J. W. Lee, "Two-sample rank tests for acceleration in cure models," *Stat. Med.* vol. 14 pp. 2111–2118, 1995.

R. Maller and X. Zhou, *Survival Analysis with Long-Term Survivors*, John Wiley: New York, 1996.

W. Nelson, "Theory and applications of hazard plotting for censored failure data," *Technometrics* vol. 14 pp. 945–965, 1972.

R. Peto and J. Peto "Asymptotically efficient rank invariant test procedures (with discussion)," *J. Roy. Stat. Soc. Ser. A* vol. 135 pp. 185–198, 1972.

D. A. Pierce, "The asymptotic effect of substituting estimators for parameters in certain types of statistics," *Ann. Stat.* vol. 10, 475–478, 1982.

S. M. Scholl, A. Fourquet, B. Asselain, J.-Y. Pierga, J. R. Vilcoq, J. C. Durand, T. Dorval, T. Palangié, M. Jouve, P. Beuzeboc, E. Garcio-Giralt, R. J. Salmon, A. De la Rochefordière, F. Campana, and P. Pouillart, "Neoadjuvant versus adjuvant chemotherapy in premenopausal patients with tumours considered as too large for breast conserving surgery: Preliminary results of a randomised trial: S6," *Eur. J. Cancer* vol. 30A pp. 645–652, 1994.

R. Sposto, H. N. Sather, and S. A. Baker, "A comparison of tests of the difference in the proportion of patients who are cured," *Biometrics* vol. 48 pp. 87–99, 1992.

J. Taylor, "Semi-parametric estimation in failure time mixture models," *Biometrics* vol. 51 pp. 899–907, 1995.

A. Tsodikov, "A proportional hazards model taking account of long-term survivors," *Biometrics* vol. 54 pp. 1508–1516, 1998.

A. Tsodikov, "Semi-parametric models of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage," *Stat. Med.* vol. 30 pp. 895–920, 2002.

A. Yakovlev and A. Tsodikov, *Stochastic Models of Tumor Latency and their Biostatistical Applications*, World Scientific: Singapore, 1996.

Time Interval to the Development of Breast Carcinoma after Treatment for Hodgkin Disease

Merideth M. M. Wendland, M.D.[1]
Alexander Tsodikov, Ph.D.[2]
Martha J. Glenn, M.D.[3]
David K. Gaffney, M.D., Ph.D.[1]

[1] Department of Radiation Oncology, The Huntsman Cancer Institute and the University of Utah, Salt Lake City, Utah.

[2] Department of Epidemiology and Preventive Medicine, Division of Biostatistics, University of California at Davis, Davis, California.

[3] Department of Medicine, The Huntsman Cancer Institute and the University of Utah, Salt Lake City, Utah.

Address for reprints: David K. Gaffney, M.D., Ph.D., Department of Radiation Oncology, University of Utah, 50 North Medical Drive, Room AB25, Salt Lake City, UT 84132; Fax: (801) 585-3502; E-mail: david.k.gaffney@hci.utah.edu

Received January 15, 2004; revision received April 21, 2004; accepted June 17, 2004.

BACKGROUND. Women with Hodgkin disease (HD) who received mantle irradiation had an increased risk of developing breast carcinoma. The authors examined the influence of radiotherapy on the time interval to the development of breast carcinoma.

METHODS. Using population, cancer incidence, and survival data from the Surveillance, Epidemiology, and End Results (SEER) registries, standardized incidence ratios (SIR) were calculated and Kaplan–Meier curves were constructed to estimate breast carcinoma-free survival in women with HD treated with and without radiotherapy. The log-rank test was utilized and multivariate proportional hazard regression analysis was performed. Multivariate analysis was also performed using the PHPH regression model.

RESULTS. In 9 SEER registries, 8036 females were identified who were diagnosed with HD between 1973 and 1999. Of these women, 183 (2.3%) were subsequently diagnosed with breast carcinoma. The use of radiotherapy in the treatment of HD resulted in an increased risk of development of breast carcinoma (SIR = 1.90, $P < 0.01$). The log-rank test and proportional hazard regression model failed to detect a difference ($P = 0.79$) in breast carcinoma-free survival for women treated with and without radiotherapy. The PHPH regression model revealed that the use of radiotherapy had an adverse effect on long-term survival (relative risk [RR] = 1.84, $P = 0.01$), but was associated with a short-term survival advantage (RR = 0.45, $P = 0.01$).

CONCLUSIONS. Use of the PHPH model indicated that the use of radiotherapy in the treatment of HD resulted in an increased long-term risk for the subsequent development of breast carcinoma, but conferred a short-term reduction. ***Cancer* 2004;101:1275–82.** © *2004 American Cancer Society.*

KEYWORDS: secondary malignancy, breast carcinoma, Hodgkin disease, radiotherapy.

Through the use of comprehensive radiotherapy and combination chemotherapy regimens, the prognosis for patients with Hodgkin disease (HD) has improved dramatically in the last several decades. Radiotherapy and alkylating chemotherapy agents are themselves carcinogenic and, unfortunately, the dramatic gains in survival for patients with HD have been accompanied by a significant increase in the risk of secondary malignancies.[1–22] The leading cause of death among 15-year survivors of HD is second cancers.[23–25] As the number of long-term survivors continues to increase, the long-term sequelae of the modalities used to treat HD become progressively more important.

An increased incidence of both hematologic and solid tumors has been observed in patients previously treated for HD.[1,3–8,10,12,13,15–20] Breast carcinoma is the most common solid tumor diagnosed in

DOI 10.1002/cncr.20516
Published online 4 August 2004 in Wiley InterScience (www.interscience.wiley.com).

female survivors of HD and the risk varies with several patient and treatment-related factors.[5,8,10,13,15,16,18–20] Recently, several groups have evaluated not only the incidence of breast carcinoma in female survivors of HD, but also the relation to the age at treatment, site treated, gender of the patient, attained age, radiotherapy dose, ovarian function, alkylator dose, and other treatment modalities used.[5,7–14,16–22,26] The current study sought to further investigate the time interval to the development of breast carcinoma in women with HD treated with and without radiotherapy. To provide an adequate model for our data, we utilized the PHPH statistical model to evaluate both the short-term and long-term effects that radiotherapy has on breast carcinoma-free survival of female survivors of HD.[27,28] The PHPH model is composed of two PH models, one for the long-term effect and another to model the short-term effect of a given intervention.

MATERIALS AND METHODS

Data were extracted from nine Surveillance, Epidemiology, and End Results (SEER) registries (http://seer.cancer.gov). The cohort consisted of 8036 women who were diagnosed with HD and received their primary treatment between 1973 and 1999. These patients were analyzed for incidence of subsequently diagnosed breast carcinoma based on whether or not they received radiotherapy for treatment of HD.

Analysis of Incidence

Analysis of incidence is designed to compare the observed number of breast carcinoma patients among females previously treated for HD and the number of observed patients in the general population. The goal of this evaluation is to determine if the incidence of breast carcinoma is higher in patients with HD than in the general population and, if so, how this alteration in incidence is affected by radiotherapy.

SEER breast carcinoma and population files were used to derive the incidence of primary breast carcinoma (number of cases per person-year) by age at diagnosis of HD and year of HD diagnosis (http://seer.cancer.gov). To derive the set at risk, counts of women alive with breast carcinoma were subtracted from female population estimates based on age and year. For each age-year cell, incidence was represented by a ratio of the count of primary breast carcinoma cases observed within the year over the average number of females at risk in the cell. For the subset of patients with HD, the expected number of breast carcinoma cases was calculated based on the methods described by Breslow et al.[29] The follow-up for each particular patient with HD was represented as a line segment on the so-called Lexis diagram (a representation of the population history on the age-by-year plane). The number of person-years was determined within each cell that intersected with the line segment. To calculate the expected number of breast carcinoma cases for the patients with HD, the number of person-years was multiplied by the incidence in the cell and summed over all cells that intersected the individual follow-up history. The sum of the individual histories for the cohort of patients with HD resulted in the expected number of breast carcinoma cases in the cohort. The standardized incidence ratio (SIR) was calculated by dividing the observed (Obs) count of breast carcinomas by the expected (Exp) count. Estimates of the expected number of cases were based on a very large general population. Therefore, in the subsequent analysis, their variability was ignored. The hypothesis that the mean SIR is equal to 1 was tested using the chi-square statistic $(Obs-Exp)^2/Exp$. Confidence intervals (CI) were derived from the Poisson distribution of the counts. For all SIR values, 95% CI were calculated. Two-sided P values were calculated and significant differences were defined as $P \leq 0.05$.

Survival Analysis

Survival analysis methodology is a more precise instrument than the analysis of incidence based on Poisson processes—it is not based on grouped data, it treats the unknown baseline rates as nuisance parameters to be estimated, and it takes explicit account of individual follow-up and censoring. Multivariate survival analysis was performed within the cohort of patients with HD. Time to the diagnosis of breast carcinoma after treatment for HD data do not support the proportional hazard (PH) model most commonly used to analyze such data.[30] To provide an adequate model for the data, we used the PHPH model, recently introduced by a number of authors to describe departures from proportionality of hazards in the presence of long-term survivors.[27] This model includes the PH model as a nested special case and has the form

$$G(t\|z) = \exp[\theta_0\theta(z)\{1 - F^{\eta(z)}(t)\}]$$

where G is a survival function, F is the baseline survival function, θ_0 is the baseline long-term survival rate, and z is the vector of explanatory variables coding the effect of radiotherapy and age on breast carcinoma-free survival. The predictors θ and η represent relative risks (RR) for long-term and short-term survival, respectively. If $\eta = 1$, then the PHPH model becomes the PH model.

The long-term effect models the chance of developing breast carcinoma after treatment for HD. The short-term effect models variations in how quickly

TABLE 1
Patient Characteristics of the 8036 Women Identified in the SEER Database Who Were Diagnosed with Hodgkin Disease between 1973–1999 Based on Treatment with or without RT

Characteristics	RT	No RT	Total	Total (%)
Age at diagnosis (yrs)				
0–19	917	479	1396	17.4
20–29	1538	844	2382	29.6
30–39	833	538	1371	17.1
≥40	754	1434	2188	27.2
Unknown			699	8.7
Total	4042	3295	8036	
Calender yr of diagnosis				
1973–1974	279	169	448	5.6
1975–1979	811	473	1284	16.0
1980–1984	806	621	1427	17.8
1985–1989	876	664	1540	19.2
1990–1994	844	826	1670	20.8
1995–1999	817	850	1667	20.7
Total	4433	3603	8036	
Stage at diagnosis				
Local	845	501	1346	16.7
Regional	1443	688	2131	26.5
Distant	476	1240	1716	21.4
Unknown			2843	35.4
Total	2764	2429	8036	
Subsequent breast carcinoma				
Yes	133	50	183	2.3
No	4300	3553	7853	97.7
Total	4433	3603	8036	
Median follow-up (yrs)	8.5	4.0		

SEER: Surveillance, Epidemiology, and End Results program; RT: radio therapy.

secondary cancers develop. We may speculate that the long-term effect is associated with the overall carcinogenic potential of HD and its treatment, whereas the short-term effect is associated with the timing and biology of breast carcinoma latency after HD. Inference procedures for the PHPH model as a member of the so-called nonlinear transformation models family are provided in Tsodikov.[27,28] If long-term and short-term effects are of the opposite sign, survival curves may cross. The PH model is insensitive to such alternatives. The responses reproduced by the PHPH model are more diverse than those modeled by the PH family. In particular, the PHPH model represents the survival curves by a superposition of the long-term and the short-term effects. If $\theta = 1$, long-term survival for any level of z will be the same, whereas the survival curves will be different in the short term.

RESULTS

The patient characteristics of the 8036 women identified in the 9 SEER registries diagnosed with HD between 1973 and 1999 are presented in Table 1. The median age at HD diagnosis was 32.0 years. Based on SEER summary staging, 16.7% of women had localized disease, 26.5% had regional disease, 21.4% had distant disease, and staging information was not available for 35.4% of patients. Women with local and regional disease were more likely to be treated with radiotherapy (62.8% and 67.7%, respectively) than women with distant disease (27.7%). Although information regarding the use of radiotherapy in initial treatment was not available for 8.7% of patients, of the remaining 7337 women, 55.1% received radiotherapy and 44.9% did not.

TABLE 2
Patient Characteristics of 183 Women Who Developed Breast Carcinoma after Treatment for HD with or without RT

Characteristics	RT	No RT	Total	Total (%)
Age at diagnosis of HD (yrs)				
0–20	39	7	46	25.1
21–30	36	8	44	24.0
31–40	28	8	36	19.7
>40	17	23	40	21.9
Unknown	13	4	17	9.3
Total	133	50	183	
Calendar yr of diagnosis of HD				
1973–1974	16	2	18	9.8
1975–1979	50	12	62	33.9
1980–1984	45	17	62	33.9
1985–1989	17	8	25	13.7
1990–1994	5	9	14	7.7
1995–1999	0	2	2	1.1
Total	133	50	183	
Stage of HD				
Local	14	11	25	13.7
Regional	15	2	17	9.3
Distant	7	10	17	9.3
Unknown	97	27	124	67.8
Total	133	50	183	
Median follow-up (yrs)	14.0	9.0		

HD: Hodgkin disease; RT: radio therapy.

One hundred eighty-three women (2.3%) were diagnosed subsequently with breast carcinoma. The characteristics of these women are displayed in Table 2. The median follow-up was 14.0 years for women with HD treated with radiotherapy and 9.0 years for women who did not receive radiotherapy as part of their initial HD treatment.

The results of the analysis of breast carcinoma incidence are presented in Table 3. Women with a history of HD, regardless of whether they received radiotherapy or not as part of their treatment, had an increased risk of breast carcinoma compared with the general population. This risk was greater for women who received radiotherapy (SIR = 3.17, $P < 0.01$) than for women who did not receive radiotherapy (SIR = 1.67, $P < 0.01$). Women with HD who were treated

TABLE 3
Analysis of Breast Carcinoma Incidence and Hypotheses Testing

Effect	Observed count (HD)	Expected count (based on general population)	SIR (95% CI)	*P* value (SIR = 1)
HD, no RT vs. general population	50.00	29.95	1.67 (1.24, 2.20)	< 0.01
HD, RT vs. general population	134.00	42.23	3.17 (2.66, 3.79)	< 0.01
HD, RT vs. HD, no RT (homogeneity)	134.00 vs. 50.00	29.95 vs. 42.23	1.90	< 0.01

HD: Hodgkin disease; general population: based on nine Surveillance, Epidemiology, and End Results program registries; SIR; standardized incidence ratio; CI: confience interval; RT: radio therapy.

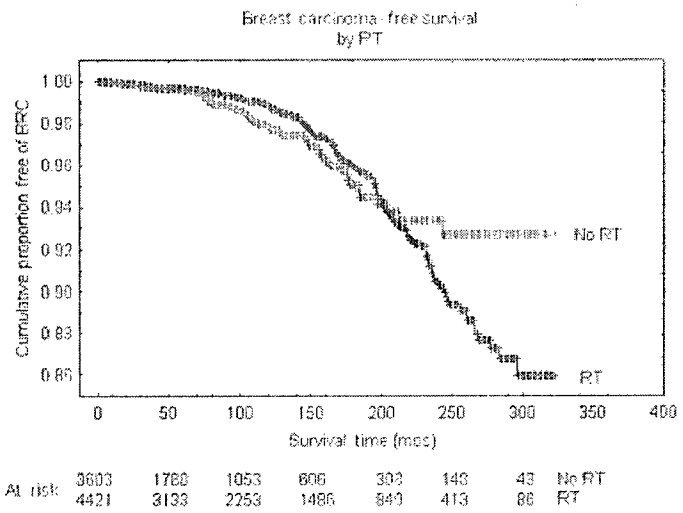


FIGURE 1. Kaplan–Meier curves for breast carcinoma-free survival from the time of diagnosis of Hodgkin disease for women treated with and without radiotherapy. BRC: breast carcinoma; RT: radiotherapy.

with radiotherapy had a greater risk of developing breast carcinoma than women who did not receive radiation (SIR = 1.90, $P < 0.01$).

To study the effect of radiotherapy in greater detail, we performed survival analyses of the time to diagnosis of breast carcinoma after treatment for HD by age at diagnosis of HD and by whether radiotherapy was received. The Kaplan–Meier curves for women treated with and without radiotherapy cross at approximately 18 years after the diagnosis of HD (Fig. 1). Figure 2 demonstrates that the expected survival curves under the PH model are virtually the same for the two groups and the log-rank test and the PH model failed to detect any difference between the groups with respect to breast carcinoma-free survival ($P = 0.79$). The search for a model that would be sensitive to the family of alternatives as shown in Figure 1 has led us to the PHPH model. Two factors were included in the model—age in years (0–19, 20–29, 30–39, ≥ 40) and the use of radiotherapy in the treatment of HD (yes or no). The observed and expected survival curves using the PHPH model are in a very good agreement (Fig. 3). The results of multivariate analysis, variable selection, and hypothesis testing based on the PHPH model are presented in Table 4. There is a significant adverse effect of the use of radiotherapy on long-term breast carcinoma-free survival (RR = 1.84, $P = 0.01$). However, the radiotherapy group enjoys a short-term reduction in breast carcinoma (RR = 0.45, $P = 0.01$). Long-term breast carcinoma-free survival was not influenced by the age at diagnosis of HD ($P = 0.18$). Conversely, short-term breast carcinoma-free survival was significantly dependent on age at HD diagnosis ($P \leq 0.01$), with an adverse effect for women > 31 years that increased with increasing age.

DISCUSSION

The increased risk for development of breast carcinoma conferred by the use of radiotherapy in the treatment of HD has been documented in many studies.[1,3,5,7–22] After evaluation of a very large population database, our independent calculation of the excess risk of developing breast carcinoma is in agreement with reported values in the literature.[7–9,11,18,20—22] Women whot received radiotherapy for treatment of HD had an SIR of 3.17 ($P < 0.01$) for breast carcinoma compared with the general population and an SIR of 1.90 ($P < 0.01$) when compared with women with HD who did not receive radiotherapy. The PH model was unable to detect a difference ($P = 0.79$) in breast carcinoma-free survival between the two groups. Us-

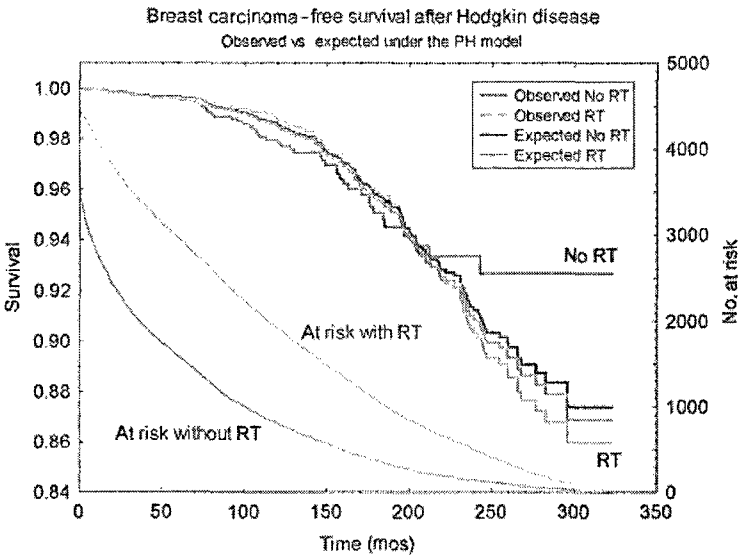


FIGURE 2. Observed versus expected breast carcinoma-free survival from the time of Hodgkin disease diagnosis for women treated with and without radiotherapy. Expected cases are predicted by use of the PH model. PH: proportional hazards model; RT: radiotherapy.

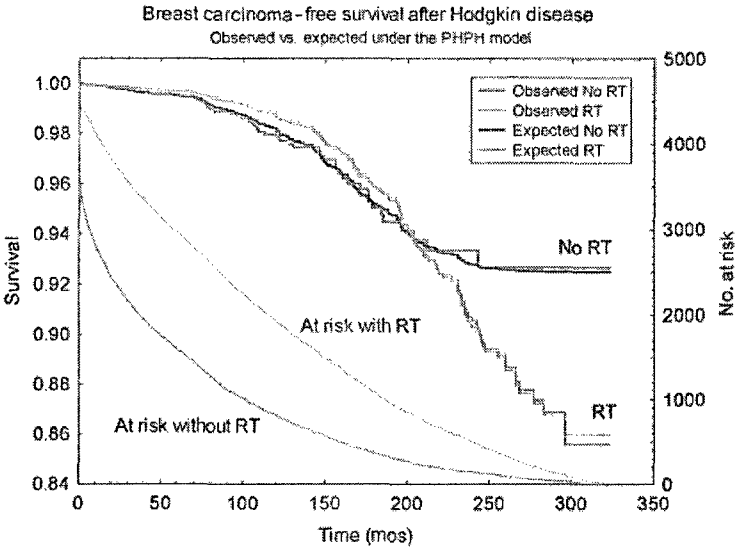


FIGURE 3. Observed versus expected breast carcinoma-free survival from the time of Hodgkin disease diagnosis for women treated with and without radiotherapy. Expected cases are predicted by use of the PHPH model. PHPH: PHPH regression model; RT: radiotherapy.

ing the PHPH model, a sophisticated statistical instrument that, unlike the PH model, has the ability to model crossing curves, a significant difference in breast carcinoma-free survival was detected. Although the use of radiotherapy in the treatment of HD results in an adverse effect on long-term breast carcinoma-free survival, it also results in a short-term reduction in the subsequent diagnosis of breast carcinoma. This noteworthy finding deserves further explanation.

Breast carcinomas observed after HD may stem from two distinct categories, i.e., those occurring spontaneously and those induced by radiotherapy. Spontaneously occurring breast carcinoma may be preexistent in a subclinical form at the time of treatment for HD, or it may develop at some point after treatment. We may assume that the time interval from HD diagnosis to the diagnosis of a preexisting breast carcinoma is shorter than for cancers that have yet to develop, as a portion of the latency period for preexisting neoplasms has already elapsed at the time of HD diagnosis. In contrast, the use of radiotherapy for the treatment of HD may induce breast carcinomas that have their full latency period ahead. As a result, the average time it takes for a clinically apparent breast carcinoma to be diagnosed is longer in the group treated with radiotherapy whereas the overall long-term risk in this group is higher as radiotherapy-induced secondary malignancies develop in addition to the number of naturally occurring spontaneous malignancies. Another possible explanation is that radio-

TABLE 4
Multivariate Survival Analysis and Hypotheses Testing[a]

Effect	RR (95% CI)	*P* value
RT vs. no RT, long term	1.84 (1.18, 2.87)	0.01
RT vs. no RT, short term	0.45 (0.26, 0.79)	0.01
Age, long term	—	0.18
Age 21–30 vs. ≤ 20, short-term	—	0.38
Age 31–40 vs. ≤ 30, short term	2.06 (1.26, 3.37)	0.01
Age > 40 vs. ≤ 30, short term	4.15 (2.68, 6.44)	< 0.01

[a] Estimates are presented for significant effects only.
RR and *P* values are adjusted for confounding using a multivariate survival model.
RR: relative risk or hazard ratio; CI: confidence interval; RT: radio therapy; age: age at treatment for Hodgkin disease.

therapy exerts a therapeutic effect on preexistent, subclinical breast carcinoma, thus eliminating breast carcinomas whose latency period has partially elapsed. The overall effect is, again, a lengthening of the average interval for the diagnosis of breast carcinoma in women with HD treated with radiotherapy compared with women who did not receive radiotherapy.

Alternatively, a positive short-term effect of radiotherapy on breast carcinoma risk may not be a causal one. One other explanation is that younger patients may have a stronger immune system and a better overall heath status than older patients. For this reason, breast carcinoma latency may be longer in younger patients. In some cohorts, radiotherapy may be applied more often in younger patients, as was observed in our analysis. As a result, breast carcinoma cells may show longer latency in the generally younger patients who receive radiotherapy.

Radiotherapy for HD, however, may lengthen the interval to the subsequent development of breast carcinoma through various mechanisms. Radiation may exert a therapeutic effect on subclinical breast carcinoma. Radiotherapy may also result in an alteration of the hormonal milieu that promotes breast carcinoma development. Travis et al.[22] demonstrated that women who received a radiotherapy dose of ≥ 5 Gray to the ovaries had a decreased risk (RR = 0.4) of subsequently developing breast carcinoma compared with women who received lower doses. Unfortunately, such detailed information, including the size, shape, and location of radiation portals and the dose to the ovaries, is not recorded in the SEER database. It is conceivable, however, that women who received radiotherapy to the abdomen and pelvis received enough of a dose to result in ovarian dysfunction, resulting in hormonal alterations and a decreased risk of secondary breast carcinoma.

Second malignancies remain an important source of morbidity and mortality for long-term survivors of HD. Comprehensive analysis concerning the malignant potential and natural histories of such cancers is not possible based on the information provided in the SEER registries and was not performed in the current study. Such important questions would require a detailed database with more rigorous follow-up information than that provided by the SEER registries.

In several other studies, the greatest risk for second primary breast carcinoma after treatment for HD has been reported for young adolescents.[5,12,17,20,31] The proposed explanation for this observation is that radiotherapy carries the greatest carcinogenic potential for actively proliferating breast tissue. Prepubescent girls and women whose mammary tissue has completed proliferation may have an increased overall risk, but not to the degree observed for adolescents. Unlike several other studies, the long-term risk of developing a second primary breast carcinoma did not vary based on age at HD diagnosis on multivariate analysis ($P = 0.18$). One explanation for the lack of effect of age at HD diagnosis on the risk of subsequent breast carcinoma in the current analysis is that the median age at diagnosis of HD was 32.0 years of age. The majority of patients (53%) were ≥ 30 years of age when HD was diagnosed and only 17.4% of patients were < 20 years of age and in what many consider to be the highest risk group.

The range of follow-up times may have an effect on the accuracy of the estimation of the long-term risk of developing breast carcinoma after treatment for HD. Differential follow-up between patients treated with and without radiotherapy is explained, in part, by the finding that in the current study, younger patients are treated with radiotherapy more often than older patients. Due to differences in expected residual survival, the group of patients treated with radiotherapy has a longer length of follow-up. The shorter follow-up for patients who did not receive radiotherapy may have resulted in a lack of power to detect differences in the long-term risk for developing breast carcinoma by age group in the current analysis. Women who received radiotherapy had a much longer median follow-up (8.5 years) than women who did not receive radiotherapy (4.0 years). This difference in follow-up was even greater for the women who were subsequently diagnosed with breast carcinoma, with a median follow-up of 14.0 years for women who received radiotherapy compared with 9.0 years for women who did not receive radiotherapy. Statisticians have formulated explicit conditions as to what kind of loss to follow-up is likely to bias the conclusions of statistical analysis. Bias would occur if loss to follow-up were associated with the time to development of breast

carcinoma (informative censoring) in either group. Unfortunately, it is impossible to verify if this is occurring based on data representing competing risks of loss to follow-up compared with the development of breast carcinoma.[32] In other words, neither descriptive statistics nor a more sophisticated statistical analysis is able to verify if data are subject to informative loss to follow-up.

Differential follow-up between the groups is not in itself a source of bias. There are several possible explanations as to why patients who receive radiotherapy have longer follow-up times. The longer follow-up in the group that received radiotherapy is reflective of the finding that radiotherapy was used alone as a curative modality in the earlier years of the current analysis and chemotherapy was used in later years. Also, in this cohort, the mean age of patients who received radiotherapy is 34 years whereas patients who did not receive radiotherapy as part of their treatment regimen were on average 45 years old at the time of diagnosis of HD. This results in a generally longer follow-up in the radiotherapy group, as follow-up is affected by differential expected residual survival. It should be stressed that the median lengths of follow-up presented in Tables 1 and 2 are based on either time to the development of breast carcinoma or loss to follow-up, whichever comes first. Therefore, the median follow-up periods for the two groups should be interpreted with caution.

There are several difficulties inherent in using the SEER registries to define a study population. These include the accuracy and completeness of the database in scoring treatment modalities used. Alkylating agents have been shown to decrease the risk of developing second primary breast carcinoma, presumably by inducing ovarian dysfunction, thus altering hormone levels and thereby influencing the development of breast carcinoma.[7,18,20–22] In two registry-based studies, the sensitivities for the receipt of systemic therapy were only 27.0% and 55.6%.[33,34] Due to the inaccuracy and incompleteness of chemotherapy reporting in the SEER registries with regards to agents used, total doses, dose intensity, and combination treatment with radiotherapy, separate analyses evaluating the role of chemotherapy in the development of second primary breast carcinoma were not performed in the current study. In addition, chemotherapy may alter the biology of second primary breast carcinoma.[7,18,21,22] The two groups in our study may be unbalanced with respect to the receipt of chemotherapy.

The accuracy of the SEER database with respect to scoring the use of radiotherapy has also been questioned. By cross-referencing treatment data recorded on Medicare reimbursement forms, one study determined that the use of radiotherapy in the treatment of breast carcinoma was not documented in the SEER database in approximately 18% of patients.[35] Therefore, information regarding the use of radiotherapy based on data gleaned from the SEER registries should be interpreted with caution.

In summary, the PH model commonly used to analyze survival data such as those presented in the current report was unable to detect a difference in breast carcinoma-free survival in women with HD treated with and without radiotherapy. This is because the survival curves cross as the hazard ratios for the two groups change over time. The PHPH regression model, which is sensitive to changes in hazard ratios and is able to analyze data in the presence of long-term survivors, provided a good model for our data. Using the PHPH model in a multivariate analysis, the use of radiotherapy in the treatment of HD is associated with a short-term reduction in the subsequent diagnosis of breast carcinoma and has an adverse effect on long-term breast carcinoma-free survival.

REFERENCES

1. Kushner BH, Zauber A, Tan CT. Second malignancies after childhood Hodgkin's disease. The Memorial Sloan-Kettering Cancer Center experience. *Cancer.* 1988;62:1364–1370.
2. Tarbell NJ, Gelber RD, Weinstein HJ, Mauch P. Sex differences in risk of second malignant tumours after Hodgkin's disease in childhood. *Lancet.* 1993;341:1428–1432.
3. Beaty O 3rd, Hudson MM, Greenwald C, et al. Subsequent malignancies in children and adolescents after treatment for Hodgkin's disease. *J Clin Oncol.* 1995;13:603–609.
4. Meadows AT, Obringer AC, Marrero O, et al. Second malignant neoplasms following childhood Hodgkin's disease: treatment and splenectomy as risk factors. *Med Pediatr Oncol.* 1989;17:477–484.
5. Sankila R, Garwicz S, Olsen JH, et al. Risk of subsequent malignant neoplasms among 1,641 Hodgkin's disease patients diagnosed in childhood and adolescence: a population-based cohort study in the five Nordic countries. Association of the Nordic Cancer Registries and the Nordic Society of Pediatric Hematology and Oncology. *J Clin Oncol.* 1996;14:1442–1446.
6. Swerdlow AJ, Douglas AJ, Hudson GV, et al. Risk of second primary cancers after Hodgkin's disease by type of treatment: analysis of 2846 patients in the British National Lymphoma Investigation. *BMJ.* 1992;304:1137–1143.
7. Swerdlow AJ, Barber JA, Hudson GV, et al. Risk of second malignancy after Hodgkin's disease in a collaborative British cohort: the relation to age at treatment. *J Clin Oncol.* 2000; 18:498–509.
8. Mauch PM, Kalish LA, Marcus KC, et al. Second malignancies after treatment for laparotomy staged IA-IIIB Hodgkin's disease: long-term analysis of risk factors and outcome. *Blood.* 1996;87:3625–3632.
9. Hancock SL, Tucker MA, Hoppe RT. Breast cancer after treatment of Hodgkin's disease. *J Natl Cancer Inst.* 1993;85: 25–31.

10. Wolden SL, Lamborn KR, Cleary SF, Tate DJ, Donaldson SS. Second cancers following pediatric Hodgkin's disease. *J Clin Oncol.* 1998;16:536–544.
11. Wolden SL, Hancock SL, Carlson RW, Goffinet DR, Jeffrey SS, Hoppe RT. Management of breast cancer after Hodgkin's disease. *J Clin Oncol.* 2000;18:765–772.
12. Bhatia S, Robison LL, Oberlin O, Greenberg, et al. Breast cancer and other second neoplasms after childhood Hodgkin's disease. *N Engl J Med.* 1996;334:745–751.
13. Green DM, Hyland A, Barcos MP, et al. Second malignant neoplasms after treatment for Hodgkin's disease in childhood or adolescence. *J Clin Oncol.* 2000;18:1492–1499.
14. Aisenberg AC, Finkelstein DM, Doppke KP, Koerner FC, Boivin JF, Willett CG. High risk of breast carcinoma after irradiation of young women with Hodgkin's disease. *Cancer.* 1997;79:1203–1210.
15. Negalia JP, Friedman DL, Yasui Y, et al. Second malignant neoplasms in five-year survivors of childhood cancer: childhood cancer survivor study. *J Natl Cancer Inst.* 2001;93:618–629.
16. Foss Abrahamsen J, Andersen A, Nome O, et al. Long-term risk of second malignancy after treatment of Hodgkin's disease: the influence of treatment, age and follow-up time. *Ann Oncol.* 2002;13:1786–1791.
17. Metayer C, Lynch CF, Clarke EA, et al. Second cancers among long-term survivors of Hodgkin's disease diagnosed in childhood and adolescence. *J Clin Oncol.* 2000;18:2435–2443.
18. Dores GM, Metayer C, Curtis RE, et al. Second malignant neoplasms among long-term survivors of Hodgkin's disease: a population-based evaluation over 25 years. *J Clin Oncol.* 2002;20:3484–3494.
19. Ng AK, Bernardo MV, Weller E, et al. Second malignancy after Hodgkin disease treated with radiation therapy with or without chemotherapy: long-term risks and risk factors. *Blood.* 2002;100:1989–1996.
20. van Leeuwen FE, Klokman WJ, van't Veer MB, et al. Long-term risk of second malignancy in survivors of Hodgkin's disease treated during adolescence or young adulthood. *J Clin Oncol.* 2000;18:487–497.
21. van Leeuwen FE, Klokman WJ, Stovall M, et al. Roles of radiation dose, chemotherapy, and hormonal factors in breast cancer following Hodgkin's disease. *J Natl Cancer Inst.* 2003;95:971–980.
22. Travis LB, Hill DA, Dores GM, et al. Breast cancer following radiotherapy and chemotherapy among young women with Hodgkin disease. *JAMA.* 2003;290:465–475.
23. Hoppe RT. Hodgkin's disease: complications of therapy and excess mortality. *Ann Oncol.* 1997;8(Suppl. 1):115–118.
24. Mauch PM, Kalish LA, Marcus KC, et al. Long-term survival in Hodgkin's disease [abstract]. *Cancer J Sci Am.* 1995;1:33.
25. Ng AK, Bernardo MP, Weller E, et al. Long-term survival and competing causes of death in patients with early-stage Hodgkin's disease treated at age 50 or younger. *J Clin Oncol.* 2002;20:2101–2108.
26. Travis LB, Curtis RE, Boice JD Jr. Late effects of treatment for childhood Hodgkin's disease. *N Engl J Med.* 1996;335:352–353.
27. Tsodikov A. Semiparametric models of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. *Stat Med.* 2002;21:895–920.
28. Tsodikov A. Semiparametric models: a generalized self-consistency approach. *J R Stat Soc [B].* 2003;65:759–774.
29. Breslow NE, Lubin JH, Marek P, Langholz B. Multiplicative models and cohort analysis. *J Am Stat Assoc.* 1983;78:1–12.
30. Cox D. Regression models and life-tables (with discussion). *J R Stat Soc [B].* 1972;34:187–220.
31. Preston DL, Mattsson A, Holmberg E, Shore R, Hildreth NG, Boice JD Jr. Radiation effects on breast cancer risk: a pooled analysis of eight cohorts. *Radiat Res.* 2002;158:220–235.
32. Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci USA.* 1975;72:20–22.
33. Bickell NA, Chassin MR. Determining the quality of breast cancer care: do tumor registries measure up? *Ann Intern Med.* 2000;132:705–710.
34. Malin JL, Kahn KL, Adams J, Kwan L, Laouri M, Ganz PA. Validity of cancer registry data for measuring the quality of breast cancer care. *J Natl Cancer Inst.* 2002;94:835–844.
35. Du X, Freeman JL, Goodwin JS. Information on radiation treatment in patients with breast cancer: the advantages of the linked Medicare and SEER data. Surveillance, Epidemiology and End Results. *J Clin Epidemiol.* 1999;52:463–470.